(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
17 January 2002 (17.01.2002)

PCT

(10) International Publication Number
WO 02/04494 A2

(51) International Patent Classification[7]: C07K 14/16

(21) International Application Number: PCT/IB01/01208

(22) International Filing Date: 9 July 2001 (09.07.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/216,995	7 July 2000 (07.07.2000)	US
2000/3437	10 July 2000 (10.07.2000)	ZA
2000/4924	15 September 2000 (15.09.2000)	ZA

(71) Applicants *(for all designated States except US)*: **MEDICAL RESEARCH COUNCIL** [ZA/ZA]; Francie van Zijl Drive, Parow Valley, 7500 Cape Town (ZA). **UNIVERSITY OF CAPE TOWN** [ZA/ZA]; Observatory, 7500 Cape Town (ZA). **UNIVERSITY OF NORTH CAROLINA AT CHAPEL HILL** [US/US]; CB 4100, Bynum Hall, Chapel Hill, NC 27599-4100 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **WILLIAMSON, Carolyn** [ZA/ZA]; University of Cape Town, Observatory, 7500 Cape Town (ZA). **SWANSTROM, Ronald, Ivar** [US/US]; University of North Carolina at Chapel Hill, CB 4100 Bynum Hall, Chapel Hill, NC 27599-4100 (US). **MORRIS, Lynn** [ZA/ZA]; National Institute for Virology, Modderfontein Road, 2131 Sandringham (ZA). **KARIM, Salim, Abdool** [ZA/ZA]; Francie van Zijl Drive, Parow Valley, 7500 Cape Town (ZA). **JOHNSTON, Robert, Edward** [US/US]; University of North Carolina at Chapel Hill, CB 4100, Bynum Hall, Chapel Hill, NC 27599-4100 (US).

(74) Agents: **CLELLAND, Sandra, Luischen** et al.; Spoor and Fisher, P.O. Box 41312, 2024 Craighall (ZA).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:

— *without international search report and to be republished upon receipt of that report*

— *entirely in electronic form (except for this front page) and available upon request from the International Bureau*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 02/04494 A2

(54) Title: PROCESS FOR THE SELECTION OF HIV-1 SUBTYPE C ISOLATES, SELECTED HIV-1 SUBTYPE ISOLATES, THEIR GENES AND MODIFICATIONS AND DERIVATIVES THEREOF

(57) Abstract: The invention provides a process for the selection of HIV-1 subtype (clade) C isolates, selected HIV-1 subtype C isolates, their genes and modifications and derivatives thereof for use in prophylactic and therapeutic vaccines to produce proteins and polypeptides for the purpose of eliciting protection against HIV infection or disease. The process for the selection of HIV subtype isolates comprises the steps of isolating viruses from recently infected subjects; generating a consensus sequence for at least part of at least one HIV gene by identifying the most common codon or amino acid among the isolated viruses; and selecting the isolated virus or viruses with a high sequence identity to the consensus sequence. HIV-1 subtype C isolates, designated Du422, Du 151 and Du 179 (assigned Accession Numbers 01032114, 00072724 and 00072725, respectively, by the European Collection of Cell Cultures) are also provided.

PROCESS FOR THE SELECTION OF HIV-1 SUBTYPE C ISOLATES, SELECTED HIV-1 SUBTYPE ISOLATES, THEIR GENES AND MODIFICATIONS AND DERIVATIVES THEREOF

BACKGROUND TO THE INVENTION

THIS invention relates to a process for the selection of HIV-1 subtype (clade) C isolates, selected HIV-1 subtype C isolates, their genes and modifications and derivatives thereof for use in prophylactic and therapeutic vaccines to produce proteins and polypeptides for the purpose of eliciting protection against HIV infection or disease.

The disease acquired immunodeficiency syndrome (AIDS) is caused by human immunodeficiency virus (HIV). Over 34 million people worldwide are thought to be living with HIV/AIDS, with over 90% of infected people living in developing countries (UNAIDS, 1999). It is estimated that 24 million infected people reside in sub-Saharan Africa and that South Africa currently has one of the world's fastest growing HIV-1 epidemics. At the end of 1999, over 22 % of pregnant women attending government antenatal clinics in South Africa were HIV positive (Department of Health, 2000). A preventative vaccine is considered to be the only feasible way to control this epidemic in the long term.

HIV shows remarkable genetic diversity that has confounded the development of a vaccine. The molecular basis of variation resides in the viral enzyme reverse transcriptase which not only introduces an error every round of replication, but also promotes recombination between viral RNAs. Based on phylogenetic analysis of sequences, HIV has been classified into a number of groups: the M (major group) which comprises subtypes A to H and K, the O (outlier group) and the N (non-M, non-O group). Recently recombinant viruses have been more frequently identified and there are a number which have spread significantly and established epidemics (circulating recombinant forms or CRF) such as subtype A/G recombinant in West Africa, and CRF A/E recombinant in Thailand (Robertson *et al.*, 2000).

Subtype C predominates in the Southern African region which includes Botswana, Zimbabwe, Zambia, Malawi, Mozambique and South Africa. In addition, increasing numbers of subtype C infections are being detected in the Southern region of Tanzania. This subtype also predominates in Ethiopia and India and is becoming more important in China.

A possible further obstacle to vaccine development is that the biological properties of HIV change as disease progresses. HIV requires two receptors to infect cells, the CD4 and co-receptors of which CCR5 and CXCR4 are the major co-receptors used by HIV-1 strains. The most commonly transmitted phenotype is non-syncytium inducing (NSI), macrophage-tropic viruses that utilise the CCR5 co-receptor for entry (R5 viruses). Langerhans cells in the mucosa are thought to selectively pick up R5 variants at the portal of entry and transport them to the lymph nodes where they undergo replication and expansion. As the infection progresses, viruses evolve that have increased replicative capacity and the ability to grow in T cell lines. These syncytium-inducing (SI) T-tropic viruses use CXCR4 in conjunction with or in preference to CCR5, and in some cases also use other minor co-receptors (Connor *et al.*, 1997, Richman & Bozzette, 1994). However HIV-1 subtype C viruses appear to be unusual in that they do not readily undergo this phenotypic switch, as R5 viruses are also predominant in patients with advanced AIDS (Bjorndal *et al.*, 1999, Peeters *et al.*, 1999, Ping *et al.*, 1999, Tscherning *et al.*, 1998, Scarlatti *et al.*, 1997).

SUMMARY OF THE INVENTION

According to one aspect of the invention a process for the selection of HIV subtype isolates for use in the development of prophylactic and therapeutic pharmaceutical composition comprises the following steps:

isolating viruses from recently infected subjects;

generating a consensus sequence for at least part of at least one HIV gene by identifying the most common codon or amino acid among the isolated viruses at each position along at least part of the gene; and

selecting the isolated virus or viruses with a high sequence identity to the consensus sequence, a phenotype which is associated with transmission for the particular HIV subtype.

The isolated virus may be of the same subtype as a likely challenge strain.

The HIV subtype is preferably HIV-1 subtype C.

For HIV-1 subtype C, the phenotype which is associated with transmission is typically a virus that utilises the CCR5 co-receptor and is non syncitium inducing (NSI).

According to another aspect of the invention an HIV-1 subtype C isolate, designated Du422 and assigned Provisional Accession Number 01032114 by the European Collection of Cell Cultures, is provided.

According to another aspect of the invention an HIV-1 subtype C isolate, designated Du151 and assigned Accession Number 00072724 by the European Collection of Cell Cultures, is provided.

According to another aspect of the invention an HIV-1 subtype C isolate, designated Du179 and assigned Accession Number 00072725 by the European Collection of Cell Cultures, is provided.

According to another aspect of the invention a molecule is provided, the molecule having:

(i) the nucleotide sequence set out in sequence as set out in Sequence I.D. No. 1;

(ii) an RNA sequence corresponding to the nucleotide sequence set out in Sequence I.D. No. 1;

(iii) a sequence which will hybridise to the nucleotide sequence set out in Sequence I.D. No. 1 or an RNA sequence corresponding to it, under strict hybridisation conditions;

(iv) a sequence which is homologous to the nucleotide sequence set out in Sequence I.D. No. 1 or an RNA sequence corresponding to it; or

(v) a sequence which is a modification or derivative of the sequence of any one of (i) to (iv).

The modified sequence is preferably that set out in Sequence I.D. No. 7.

According to another aspect of the invention a molecule is provided, the molecule having:

(i) the nucleotide sequence set out in Sequence I.D. No. 3;

(ii) an RNA sequence corresponding to the nucleotide sequence set out in Sequence I.D. No. 3;

(iii) a sequence which will hybridise to the nucleotide sequence set out in Sequence I.D. No. 3 or an RNA sequence corresponding to it, under strict hybridisation conditions;

(iv) a sequence which is homologous to the nucleotide sequence set out in Sequence I.D. No. 3 or an RNA sequence corresponding to it; or

(v) a sequence which is a modification or derivative of the sequence of any one of (i) to (iv).

The modified sequence is preferably that set out in Sequence I.D. No. 9.

According to another aspect of the invention a molecule is provided, the molecule having:

(i) the nucleotide sequence set out in Sequence I.D. No. 5;

(ii) an RNA sequence corresponding to the nucleotide sequence set out in Sequence I.D. No. 5;

(iii) a sequence which will hybridise to the nucleotide sequence set out in Sequence I.D. No. 5 or an RNA sequence corresponding to it, under strict hybridisation conditions;

(iv) a sequence which is homologous to the nucleotide sequence set out in Sequence I.D. No. 5 or an RNA sequence corresponding to it; or

(v) a sequence which is a modification or derivative of the sequence of any one of (i) to (iv).

The modified sequence is preferably that set out in Sequence I.D. No. 11.

According to another aspect of the invention a molecule is provided, the molecule having:

(i) the nucleotide sequence set out in Sequence I.D. No. 13;

(ii) an RNA sequence corresponding to the nucleotide sequence set out in Sequence I.D. No. 13;

(iii) a sequence which will hybridise to the nucleotide sequence set out in Sequence I.D. No. 13 or an RNA sequence corresponding to it, under strict hybridisation conditions;

(iv) a sequence which is homologous to the nucleotide sequence set out in Sequence I.D. No. 13 or an RNA sequence corresponding to it; or

(v) a sequence which is a modification or derivative of the sequence of any one of (i) to (iv).

The modified sequence preferably has similar or the same modifications as those set out in Sequence I.D. No. 11 for the *env* gene of the isolate Du151.

According to another aspect of the invention a polypeptide is provided, the polypeptide having:

(i) the amino acid sequence set out in Sequence I.D. No. 2; or

(ii) a sequence which is a modification or derivative of the amino acid sequence set out in Sequence I.D. No. 2.

The modified sequence is preferably that set out in Sequence I.D. No. 8.

According to another aspect of the invention a polypeptide is provided, the polypeptide having:

(i) the amino acid sequence set out in Sequence I.D. No. 4; or

(ii) a sequence which is a modification or derivative of the amino acid sequence set out in Sequence I.D. No. 4.

The modified sequence is preferably that set out in Sequence I.D. No. 10.

According to another aspect of the invention a polypeptide is provided, the polypeptide having:

(i) the amino acid sequence set out in Sequence I.D. No. 6; or

(ii) a sequence which is a modification or derivative of the amino acid sequence set out in Sequence I.D. No. 6.

The modified sequence is preferably that set out in Sequence I.D. No. 12.

According to another aspect of the invention a polypeptide is provided, the polypeptide having:

(i) the amino acid sequence set out in Sequence I.D. No. 14;

(ii) a sequence which is a modification or derivative of the amino acid sequence set out in Sequence I.D. No. 14.

The modified sequence preferably has similar or the same modifications as those set out in Sequence I.D. No. 12 for the amino acid sequence of the *env* gene of the isolate Du151.

According to another aspect of the invention a consensus amino acid sequence for the partial *gag* gene of HIV-1 subtype C is the following:

GEKLDKWEKI	RLRPGGKKHY	MLKHLVWASR	ELERFALNPG	LLETSEGCKQ [50]
IMKQLQPALQ	TGTEELRSLY	NTVATLYCVH	EKIEVRDTKE	ALDKIEEEQN [100]
KSQQ-CQQKT	QQAKAADGG-	KVSQNYPIVQ	NLQGQMVHQA	ISPRTLNAWV [150]
EEKAFSP	EVIPMFTALS	EGATPQDLNT	MLNTVGGHQA	AMQMLKDTIN [200]
EEAAEWDRLH	PVHAGPIAPG	QMREPRGSDI	AGTTSTLQEQ	IAWMTSNPPI [250]
PVGDIYKRWI	ILGLNKIVRM	YSPVSILDIK	QGPKEPFRDY	VDRFFKTLRA [300]
EQATQDVKNW	MTD [313]			

According to another aspect of the invention a consensus amino acid sequence for the partial *pol* gene of HIV-1 subtype C is the following:

LTEEKIKALT	AICEEMEKEG	KITKIGPENP	YNTPVFAIKK	KDSTKWRKL- [50]
VDFRELNKRT	QDFWEVQLGI	PHPAGLKKKK	SVTVLDVGDA	YFSVPLDEGF [100]
RKYTAFTIPS	INNETPGIRY	QYNVLPQGWK	GSPAIFQSSM	TKILEPFRAK [150]
NPEIVIYQYM	DDLYVGSDLE	IGQHRAKIEE	LREHLLKWGF	TTPDKKHQKE [200]
PPFLWMGYEL	HPDKWTVQPI	QLPEKDSWTV	NDIQKLVGKL	NWASQIYPGI [250]
KVRQLCKLLR	GAKALTDIVP	LTEEAELE [278]		

According to another aspect of the invention a consensus amino acid sequence for the partial *env* gene of HIV-1 subtype C is the following:

YCAPAGYAIL	KCNNKTFNGT	GPCNNVSTVQ	CTHGIKPVVS	TQLLLNGSLA [50]
EEEIIIRSEN	LTNNAKTIIV	HLNESVEIVC	TRPNNNTRKS	IRIGPGQTFY [100]
ATGDIIGDIR	QAHCNISEGK	WNKTLQKVKK	KLKEELYKYK	VVEIKPLGIA [150]
PTEAKRRVVE	REKRAVGIGA	VFLGFLGAAG	STMGAASITL	TVQARQLLSG [200]
IVQQQSNLLR	AIEAQQHMLQ	LTVWGIKQL [229]		

DESCRIPTION OF THE DRAWINGS

Figure 1 shows a schematic representation of the HIV-1 genome and illustrates the location of overlapping fragments that were sequenced having been generated by reverse transcriptase followed by polymerase chain reaction, in order to generate the South African consensus sequence;

Figure 2 shows a phylogenetic tree of nucleic acid sequences of various HIV-1 subtype C isolates based on the (partial) sequences of the *gag* gene of the various isolates and includes a number of consensus sequences as well as the South African consensus sequence of the present invention and a selected isolate, Du422, of the present invention;

Figure 3 shows a phylogenetic tree of nucleic acid sequences of various HIV-1 subtype C isolates based on the (partial) sequences of the *pol* gene of the various isolates and includes a number of consensus sequences as well as the South African consensus sequence of the present invention and a selected isolate, Du151, of the present invention;

Figure 4 shows a phylogenetic tree of nucleic acid sequences of various HIV-1 subtype C isolates based on the (partial) sequences of the *env* gene of the various isolates and includes a number of consensus sequences as well as the South African consensus sequence of the present invention and a selected isolate, Du151, of the present invention

Figure 5 shows how the sequences of the *gag* genes of each of a number of isolates varies from the South African consensus sequence for the *gag* gene which was developed according to the present invention;

Figure 6 shows how the sequences of the *pol* genes of each of a number of isolates varies from the South African consensus sequence for the *pol* gene which was developed according to the present invention;

Figure 7 shows how the sequences of the *env* genes of each of a number of isolates varies from the South African consensus sequence for the *env* gene which was developed according to the present invention;

Figure 8 shows a phylogenetic tree of amino acid sequences of various HIV-1 subtype C isolates based on the sequences of the (partial) *gag* gene of the various isolates and includes a number of consensus sequences as well as the South African consensus sequence of the present invention and a selected isolate, Du422, of the present invention;

Figure 9 shows a phylogenetic tree of amino acid sequences of various HIV-1 subtype C isolates based on the sequences of the (partial) *pol* gene of the various isolates and includes a Cpol consensus sequence as well as a South African consensus sequence of the present invention and a selected isolate, Du151, of the present invention;

Figure 10 shows a phylogenetic tree of amino acid sequences of various HIV-1 subtype C isolates based on the sequences of the (partial) *env* gene of the various isolates and includes a Cenv consensus sequence as well as a South African consensus sequence of the present invention and a selected isolate, Du151, of the present invention;

Figure 11 shows the percentage amino acid sequence identity of the sequenced *gag* genes of the various isolates in relation to one another, to the *gag* clone and to the South African consensus sequence for the *gag* gene and is based on a pairwise comparison of the *gag* genes of the isolates;

Figure 12 shows the percentage amino acid sequence identity of the sequenced *pol* genes of the various isolates in relation to one another, to the *pol* clone and to the South African consensus sequence for the *pol* gene and is based on a pairwise comparison of the *pol* genes of the isolates;

Figure 13 shows the percentage amino acid sequence identity of the sequenced *env* genes of the various isolates in relation to one another, to the *env* clone and to the South African consensus sequence for the *env* gene and is based on a pairwise comparison of the *env* genes of the isolates;

Figure 14 shows a phylogenetic tree analysis of nucleic acid sequences of various HIV-1 subtype C isolates (or vaccine strains) based on the

complete sequences of the *gag* genes of the various isolates and shows the *gag* gene from a selected isolate, Du422, of the present invention compared to the other subtype C sequences;

Figure 15 shows a phylogenetic tree analysis of nucleic acid sequences of various HIV-1 subtype C isolates (or vaccine strains) based on the complete sequences of the *pol* genes of the various isolates and shows the *pol* gene from a selected isolate, Du151, of the present invention compared to the other subtype C sequences;

Figure 16 shows a phylogenetic tree analysis of nucleic acid sequences of various HIV-1 subtype C isolates (or vaccine strains) based on the complete sequences of the *env* gene of the various isolates and shows the *env* gene from a selected isolate, Du151, of the present invention compared to the other subtype C sequences; and

LIST OF SEQUENCES

Sequence I.D. No 1 shows the nucleic acid sequence (cDNA) of the sequenced *gag* gene of the isolate Du422;

Sequence I.D. No 2 shows the amino acid sequence of the sequenced *gag* gene of the isolate Du422, derived from the nucleic acid sequence;

Sequence I.D. No 3 shows the nucleic acid sequence (cDNA) of the sequenced *pol* gene of the isolate Du151;

Sequence I.D. No 4 shows the amino acid sequence of the sequenced *pol* gene of the isolate Du151, derived from the nucleic acid sequence;

Sequence I.D. No 5 shows the nucleic acid sequence (cDNA) of the sequenced *env* gene of the isolate Du151;

Sequence I.D. No 6 shows the amino acid sequence of the sequenced *env* gene of the isolate Du151, derived from the nucleic acid sequence;

Sequence I.D. No 7 shows the nucleic acid sequence (DNA) of the resynthesized sequenced *gag* gene of the isolate Du422 modified to reflect human codon usage for the purposes of increased expression;

Sequence I.D. No 8 shows the amino acid sequence of the resynthesized sequenced *gag* gene of the isolate Du422 modified to reflect human codon usage for the purposes of increased expression;

Sequence I.D. No 9 shows the nucleic acid sequence (DNA) of the resynthesized sequenced *pol* gene of the isolate Du151 modified to reflect human codon usage for the purposes of increased expression;

Sequence I.D. No 10 shows the amino acid sequence of the resynthesized sequenced *pol* gene of the isolate Du151 modified to reflect human codon usage for the purposes of increased expression;

Sequence I.D. No 11 shows the nucleic acid sequence (DNA) of the resynthesized sequenced *env* gene of the isolate Du151 modified to reflect human codon usage for the purposes of increased expression;

Sequence I.D. No 12 shows the amino acid sequence of the resynthesized sequenced *env* gene of the isolate Du151 modified to reflect human codon usage for the purposes of increased expression;

Sequence I.D. No 13 shows the nucleic acid sequence (cDNA) of the sequenced *env* gene of the isolate Du179; and

Sequence I.D. No 14 shows the amino acid sequence of the sequenced *env* gene of the isolate Du179.

DETAILED DESCRIPTION OF THE INVENTION

This invention relates to the selection of HIV-1 subtype isolates and the use of their genes and modifications and derivatives thereof in making prophylactic and therapeutic pharmaceutical compositions and formulations, and in particular vaccines against HIV-1 subtype C. The compositions could therefore be used either prophylactically to prevent infection or therapeutically to prevent or modify disease. A number of factors must be taken into consideration in the development of an HIV vaccine and one aspect of the present invention relates to a process for the selection of suitable HIV isolates for the development of a vaccine.

The applicant envisages that the vaccine developed according to the above method could be used against one or more HIV subtypes other than HIV-1 subtype C.

An HIV vaccine aims to elicit both a CD8+ cytotoxic T lymphocyte (CTL) immune response as well as a neutralizing antibody response. Many current vaccine approaches have primarily focused on inducing a CTL response. It is thought that the CTL response may be more important as it is associated with the initial control of viral replication after infection, as well as control of replication during disease, and is inversely correlated with disease progression (Koup *et al.*, 1994, Ogg *et al.*, 1999 Schmitz *et al.*, 1999). The importance of CTL in protecting individuals from infection is demonstrated by their presence in highly exposed seronegative individuals such as sex-workers (Rowland-Jones *et al.*, 1998).

Knowledge of genetic diversity is highly relevant to the design of vaccines aiming at eliciting a cytotoxic T-lymphocyte (CTL) response. There are many CTL epitopes in common between viruses, particularly in the *gag* and *pol* region of the genome (HIV Molecular Immunology Database, 1998). In addition, several studies have now shown that there is a cross-reactive CTL response: individuals vaccinated with a subtype B-based vaccine could lyse autologous targets infected with a diverse group of isolates (Ferrari *et al.*, 1997); and CTLs from non-B infected individuals could lyse subtype B-primed targets (Betts *et al.* 1997; Durali *et al*, 1998). A comparison of CTL epitopes in the HIV-1 sequence database shows about 40% of gp41 and 84% of p24 epitopes are

identical or have only one amino acid difference between subtypes. Although this is a very crude analysis and does not take into consideration populations or dominant responses to certain epitopes, it does however indicate that there is a greater conservation of cytotoxic T epitopes within a subtype compared to between subtypes and that there will be a greater chance of a CTL response if the challenge virus is the same subtype as the vaccine strain.

The importance of genetic diversity in inducing a neutralizing antibody response appears to be less crucial. In general, neutralization serotypes are not related to genetic subtype. Some individuals elicit antibodies that can neutralize a broad range of viruses, including viruses of different subtypes while others fail to elicit effective neutralizing antibodies at all (Wyatt and Sodroski, 1998; Kostrikis *et al.*, 1996; Moore *et al.*, 1996). As neutralizing antibodies are largely evoked against functional domains of the virus which are essentially conserved, it is probable that HIV-1 genetic diversity may not be relevant in producing a vaccine designed to elicit neutralizing antibodies.

Viral strains used in the design of a vaccine need to be shown by genotypic analysis to be representative of the circulating strains and not an unusual or outlier strain. In addition, it is important that a vaccine strain also has the phenotype of a recently transmitted virus, which is NSI and uses the CCR5 co-receptor.

A process was developed to identify appropriate strains for use in developing a vaccine for HIV-1 subtype C. Viral isolates from acutely infected individuals were collected. They were sequenced in the *env*, *gag* and *pol* regions and the amino acid sequences for the *env*, *gag* and *pol* genes from these isolates were compared. A consensus sequence, the South African consensus sequence, was then formed by selecting the most frequently appearing amino acid at each position. The consensus sequence for each of the *gag*, *pol* and *env* genes of HIV-1 subtype C also forms an aspect of the invention. Appropriate strains for vaccine development were then selected from these isolates by comparing them with the consensus sequence and characterising them phenotypically. The isolates also form an aspect of the invention.

In order to select for NSI strains which use the CCR5 co-receptor, a well established sex worker cohort was used to identify the appropriate strains. Appropriate strains

were identified from acutely infected individuals by comparing them with the consensus sequence which had been formed. Viral isolates from fifteen acutely infected individuals were sequenced in the *env*, *gag* and *pol* and phenotypically characterised. These sequences were compared with viral isolates from fifteen asymptomatic individuals from another region having more than 500 CD4 cells and other published subtype C sequences located in the Los Alamos Database (http://ww.hiv-web.lanl.gov/).

Three potential vaccine strains, designated Du151, Du422 and Du179, were selected. Du 151 and Du 422 were selected based on amino acid homology to the consensus sequence in all three gene regions *env*, *gag* and *pol*, CCR5 tropism and ability to grow and replicate in tissue culture. Du 179 is a R5X4 virus and was selected because the patient in which this strain was found showed a high level of neutralising antibodies. The nucleotide and amino acid sequences of the three gene regions of the three isolates and modifications and derivatives thereof also form aspects of the invention.

The vaccines of the invention will be formulated in a number of different ways using a variety of different vectors. They involve encapsulating RNA or transcribed DNA sequences from the viruses in a variety of different vectors. The vaccines will contain at least part of the *gag* gene from the Du422 isolate, and at least part of the *pol* and *env* genes from the Du151 isolate of the present invention and/or at least part of the *env* gene from the Du179 isolate of the present invention or derivatives or modifications thereof.

Genes for use in DNA vaccines have been resynthesized to reflect human codon usage. The *gag* Du422 gene was designed so that the myristylation site and inhibitory sequences were removed. Similarly resynthesized gp 160 (the complete *env* gene consisting of gp 120 and gp 41) and *pol* genes will be expressed by DNA vaccines. The gp160 gene sequence has also been changed as described above for the *gag* gene to reflect human codon usage and the rev responsive element removed. The protease, inactivated reverse transcriptase and start of the RNAse H genes from Du151 *pol* are optimised for increased expression and will be joined with *gag* at an inserted Bgl1 site. The *gag-pol* frameshift will be maintained to keep the natural balance of *gag* to *pol* protein expression.

Another vaccine will contain DNA transcribed from the RNA for the *gag* gene from the Du422 isolate and RNA from the *pol* and *env* genes from the Du151 isolate and/or RNA from the *env* gene from the Du179 isolate. These genes could also be expressed as oligomeric envelope glycoprotein complexes (Progenics, USA) as published in J Virol 2000 Jan;74(2):627-43 (Binley, J.L. et al.), the adeno associated virus (AAV) (Target Genetics) and the Venezuelan equine encephalitus virus (US patent application USSN 60/216,995, which is incorporated herein by reference).

The isolation and selection of viral strains for the design of a vaccine

The following criteria were used to select appropriate strains for inclusion into HIV-1 vaccines for Southern Africa:

that the strains be genotypically representative of circulating strains;

that the strain not be an outlier strain;

that the strain be as close as possible to the consensus amino acid sequence developed according to the invention for the *env*, *gag* and *pol* genes of HIV-1 subtype C;

that the strain have an R5 phenotype, i.e. a phenotype associated with transmission for selection of the RNA or cDNA to be included for the *env* region; and

that the vaccine be able to be grown in tissue culture.

The following procedure was followed in the selection of viral strains for the design of a vaccine. A well-established sex worker cohort in Kwazulu Natal, South Africa was used to identify the appropriate strains for use in an HIV vaccine. Viral isolates from 15 acutely infected individuals were sequenced in *env*, *gag* and *pol* and were also isolated and phenotypically characterised. These sequences were compared with a similar

collection from asymptomatic individuals from the Gauteng region in South Africa as well as other published subtype C sequences.

Patients

Individuals with HIV infection were recruited from 4 regions in South Africa. Blood samples were obtained from recently infected sex workers from Kwazulu-Natal (n=13). Recent infection was defined as individuals who were previously seronegative and had became seropositive within the previous year. Samples were also collected from individuals attending out-patients clinics in Cape Town (n=2), women attending ante-natal clinics in Johannesburg (n=7) and men attending a STD clinic on a gold mine outside Johannesburg (n=8). The latter 2 groups were clinically stable and were classified as asymptomatic infections. Blood samples were collected in EDTA and used to determine the CD4 T cell count and genetic analysis of the virus. In the case of recent infections a branched chain (bDNA) assay (Chiron) to measure plasma viral load was done, and the virus was isolated. HIV-1 serostatus was determined by ELISA. The results of the CD4 T cell counts and the viral loads on the sex workers were established and information on the clinical status as at date of seroconversion, CD4, and data on the co-receptor usage is set out in Table 1 below.

Virus isolation

HIV was isolated from peripheral blood mononuclear cells (PBMC) using standard co-culture techniques with mitogen-activated donor PBMC. $2x10^6$ patient PBMC were co-cultured with $2x10^6$ donor PBMC in 12 well plates with 2 ml RPMI 1640 with 20% FCS, antibiotics and 5% IL-2 (Boehringer). Cultures were replenished twice weekly with fresh medium containing IL-2 and once with $5x10^5$/ml donor PBMC. Virus growth was monitored weekly using a commercial p24 antigen assay (Coulter). Antigen positive cultures were expanded and cultured for a further 2 weeks to obtain 40 mls of virus containing supernatant which was stored at -70^oC until use. The results of the

isolation of the viruses from the commercial sex workers is also shown in Table 1 below.

Viral phenotypes

Virus-containing supernatant was used to assess the biological phenotype of viral isolates on MT-2 and co-receptor transfected cell lines. For the MT-2 assay, 500 ul of supernatant was incubated with $5x10^4$ MT-2 cells in PRMI plus 10% FCS and antibiotics. Cultures were monitored daily for syncitia formation over 6 days. U87.CD4 cell expressing either the CCR5 or CXCR4 co-receptor were grown in DMEM with 10% FCS, antibiotics, 500 ug/ml G418 and 1 ug/ml puromycin . GHOST cells expressing minor co-receptors were grown in DMEM with 10% FCS, 500 ug/ml G418, 1 ug/ml puromycin and 100 ug/ml hygromycin. Cell lines were passaged twice weekly by trypsination. Co-receptor assays were done in 12 well plates; $5x10^4$ cells were plated in each well and allowed to adhere overnight. The following day 500ul of virus containing supernatant was added and incubated overnight to allow viral attachment and infection and washed three times the following day. Cultures were monitored on days 4, 8 and 12 for syncitia formation and p24 antigen production. Cultures that showed evidence of syncitia and increasing concentrations of p24 antigen were considered positive for viral growth. The results of co-receptor usage of the viruses from the commercial sex workers is also shown in Table 1.

TABLE 1 – COHORT OF ACUTE INFECTIONS FOR SELECTION OF VACCINE CANDIDATES

Sample ID	Sero date	Sample date	Duration of infection	CD4 count	Viral load	Co-culture p24 pos	MT-2 assay	Biotype
Du115	15 May 98	20 May 99	1 year	437*	7,597*	-	No isolate	-
Du123	17 Aug 98	17 Nov 98	3 mon	841	19,331	d6 (50pg)	NSI	R5
Du151	12 Oct 98	24 Nov 98	1.5 mon	367	>500.000	d6 (>1ng)	NSI	R5
Du156	16 Nov 98	17 Nov 98	<1 mon	404	22.122	d6 (>1ng)	NSI	R5
Du172	16 Oct 98	17 Nov 98	1 mon	793	1.916	d6 (<50pg)	NSI	R5
Du174	6 Oct 97	25 May 99	19.5 mon	634*	9,454*	d14 (>1ng)	NSI	R5
Du179	13 Aug 97	20 May 99	21 mon	394*	1,359*	d7 (<50pg)	SI	R5x4
Du204	20 May 98	20 May 99	1 year	633*	8.734*	d7 (<50pg)	NSI	R5
Du258	3 June 98	22 Jun 99	1 year	433*	9,114*	-	No isolate	-
Du281	24 July 98	17 Nov 98	4 mon	594	24.689	d6 (1ng)	NSI	R5
Du285	2 Oct 98	-	-	560*	161*	-	No isolate	-
Du368	8 Apr 98	24 Nov 98	7.5 mon	670	13,993	d6 (300pg)	NSI	R5
Du422	2 Oct 98	28 Jan 99	4 mon	397	17,118*	d6 (600 pg)	NSI	R5
Du457	17 Aug 98	17 Nov 98	3 mon	665	6,658	-	No isolate	-
Du467	26 Aug 98	-	-	671	19.268	-	No isolate	-

* date from Nov 98

Sequencing

RNA was isolated from plasma and the gene fragments were amplified from RNA using reverse transcriptase to generate a cDNA followed by PCR to generate amplified DNA segments. The positions of the PCR primers are as follows, with the second of each primer pair being used as the reverse transcriptase primer in the cDNA synthesis step (numbering using the HIV-1 HXBr sequence): *gag*1 (790-813, 1282-1303), *gag*2 (1232-1253 , 1797-1820), *pol*1 (2546-2573 , 3012-3041), *pol*2(2932-2957 , 3492-3515), *env*1 (6815-6838, 7322-7349), *env*2 (7626-7653 , 7963-7986). The amplified DNA fragments were purified using the QIAQUICK PCR Purification Kit (Qiagen, Germany). The DNA fragments were then sequenced using the upstream PCR primers as sequencing primers. Sequencing was done using the Sanger dideoxyterminator strategy with fluorescent dyes attached to the dideoxynucleotides. The sequence determination was made by electrophoresis using an ABI 377 Sequencer. A mapped illustration of an HIV-1 proviral genome showing the *pol, gag* and *env* regions sequenced as described above, is shown in Figure 1. The following regions were sequenced (numbering according to HXBr, Los Alamos database); 813 – 1282 (gag1); 1253 – 1797 (gag2); 2583 – 3012 (pol1); 2957-3515 (pol2); 6938 – 7322 (env1); 7653 – 7963 (env2), as illustrated in Figure 1.

Genotypic characterisation

To select the vaccine isolate or isolates, a survey covering portions of the three major HIV genes *gag* (313 contiguous codons, 939 bases), *pol* (278 contiguous codons, 834 bases) and *env* (229 codons in two noncontigous segments, 687 bases) was done (Figure 1). The map of Figure 1 shows the 5'long terminal repeat, the structural and functional genes (*gag*, *pol* and *env*) as well as the regulatory and accessory proteins (*vif*, *tat*, *rev*, *nef*, *vpr* and *vpu*). The gag open reading frame illustrates the regions encoding p17 matrix protein and the p24 core protein and the p7 and p6 nuclearcapsid proteins. The pol open reading frame illustrates the protease (PR) p15, reverse transcriptase (RT) p66 and the Rnase H integrase p51. The env open reading frame indicates the region coding for gp120 and the region coding for gp41.

Of a total of 31 isolates, 14 were from the Durban cohort (DU), 15 were from Johannesburg (GG and RB) and 2 from Cape Town (CT). Of these 30 were sequenced in the *gag* region, 26 in the *pol* region and 27 in the *env* region. The isolates that were sequenced are shown in Table 2.

TABLE 2 – LIST OF ISOLATES AND THE REGIONS GENES SEQUENCED

Isolate	*Gag* sequence	*Pol* sequence	*Env* sequence
CTSC1	✔	✔	-
CTSC2	✔	✔	-
DU115	✔	✔	✔
DU123	✔	-	✔
DU151	-	✔	✔
DU156	✔	✔	✔
DU172	✔	✔	✔
DU174	✔	✔	✔
DU179	✔	✔	✔
DU204	✔	✔	✔
DU258	✔	✔	✔
DU281	✔	-	✔
DU368	✔	✔	✔
DU422	✔	✔	✔
DU457	✔	✔	✔
DU467	✔	-	✔
GG1	✔	-	-
GG10	✔	✔	✔
GG3	✔	✔	✔
GG4	✔	✔	✔
GG5	✔	✔	✔
GG6	✔	✔	✔
RB12	✔	-	✔
RB13	✔	✔	✔
RB14	✔	✔	✔

RB15	✔	✔	-
RB18	✔	✔	✔
RB21	✔	✔	✔
RB22	✔	✔	✔
RB27	✔	✔	✔
RB28	✔	✔	✔

The nucleic acid sequences from the Durban (DU) Johannesburg (GG, RB) and Cape Town (CT) cohorts were phylogenetically compared to all available published subtype C sequences (obtained from the Los Alamos HIV Sequence Database) including sequences from the other southern African countries and the overall subtype C consensus from the Los Alamos HIV sequence database. This comparison was done to ensure that the selected vaccine isolates were not phylogenetic outliers when compared to the Southern African sequences and the results of the comparison are shown in Figure 2, Figure 3 and Figure 4. Figures 2 to 4 illustrate that the sequences from Southern Africa are divergent and that the Indian sequences form a separate distinct cluster from these African sequences. The South African sequences are not unique and, in general, are as related to each other as they are to other sequences from Southern Africa. Overall this suggests Indian sequences are unique from Southern African subtype C sequences and that we do not have a clonal epidemic in South Africa, but rather South African viruses reflect the diversity of subtype C viruses in the Southern African region

Determination of a consensus sequence

Amino acid sequences were derived from the sequences shown in Table 2 and were used to determine a South African consensus sequence. The most frequently appearing amino acid at each position was selected as the consensus amino acid at that position. In this way, the consensus sequence was determined along the linear length of each of the sequenced gene fragments (*gag*, *pol* and *env* gene fragments). The alignments were done using the Genetics Computer Group (GCG) programs (Pileup and Pretty), which generates a consensus sequence in this manner. These

resulted in the consensus sequence for each gene region. The alignments of the amino acid sequences and the resulting consensus sequences are shown in Figures 5, 6 and 7.

The phylogenetic tree of amino acids showing a comparison of the South African sequences is set out in Figures 8, 9 and 10. The ES2 *gag* S, which is the sequence of the cloned Du422 *gag* gene, Du151 *pol* (clone number) 8, which is the sequence of the cloned Du151 *pol* gene, and Du151 *env* (clone number) 25, which is the sequence of the cloned Du151 *env* gene, are vaccine clones. It can be seen from Figures 8, 9 and 10 that they are the same as the original isolates. These phylogenetic trees compare the relationship between the HIV proteins. South African isolates were compared with subtype A, B, C and D consensus sequences as well as with the South African consensus (Sagagcon) derived from the South African sequences, a Malawian consensus (Malgagcon) derived from Malawian sequences and overall consensuses (Cgagcon, Cpolcon and Cenvcon) derived from all subtype C sequences on the Los Alamos database.

The final choice of which isolate or isolates to use was based on the similarity of the sequence of the *gag*, *env* and *pol* genes of a particular isolate to the South African consensus sequence which had been derived as set out above as well as the availability of an R5 isolate which had good replication kinetics as shown in Table 1.

Selection of Vaccine Isolates

Based on the considerations and methodology set out above, three strains were selected from the acute infection cohort as the vaccine strains. The first strain is Du 422 for the *gag* gene, the second strain is Du151 for the *pol* and *env* genes and the third strain is Du179 which is a possible alternative for the *env* gene. These three strains were selected for the following reasons.

1. At the time the samples were obtained, Du151 had been infected for 6 weeks and had a CD4 count of 367 cells per ul of blood and a viral load above 500,000 copies per ml of plasma. Given the high viral load, and the recorded

time from infection, it is probable that the individual was still in the initial stages of viraemia prior to control of HIV replication by the immune system.

2. At the time the samples were obtained, Du422 had been infected for 4 months with a CD4 count of 397 cells per ul of blood and a viral load of 17,118 copies per ml of plasma. In contrast to Du151, this individual had already brought viral replication under control to a certain extent.

3. At the time the samples were obtained, Du179 had been infected for 21 months with a CD4 count of 394 cells per ul of blood and a viral load of 1,359 copies per ml of plasma.

Based on the analysis of the phylogenetic tree shown in Figure 8 showing the relationship between full length gp120 sequence and other isolates, and the amino acid pairwise comparison shown in Figure 11, the Du422 *gag* sequence was shown to be most similar to the South African consensus sequence shown in Figures 2 and 5. It shared 98% amino acid sequence identity with the consensus sequence. In addition, the average pairwise distance, which is the percentage difference between the DNA sequences, between the DU422 *gag* sequence and the other sequences from the seroconverters was the highest of any sequence derived from this cohort, at 93.5%, and nearly as high as the average distance of the isolates to the SA consensus sequence (94.2%). The Du422 gag gene was cloned and the specific clone gave values very similar to the original isolate: having a pairwise identity value with the SA consensus of (98%) and nearly as high an average identity value with the other isolates as the DU422 isolate (93.3%). Thus, both the original DU422 isolate sequence and the generated clone had the highest pairwise percentage similarity to other isolates with the minimal values all being above 90%.

The *pol* sequences showed the highest values for the pairwise comparisons.. Based on the analysis of the phylogenetic tree shown in Figure 9 and the pairwise identity score with the SA consensus (98.9%) shown in Figure 12, we chose the DU151 isolate as the source of the *pol* gene. Other contributing factors in this decision were that this is the same isolate that was chosen for the source of the *env* gene and that this was an isolate with excellent growth properties *in vitro*. The actual *pol* gene clone from the

DU151 isolate was somewhat more divergent from the SA consensus sequence (97.8%), and had a smaller average identity score when compared to the other isolates (95.1%). However, we judged the small increase in distance from the consensus not to be significant in this otherwise well conserved HIV-1 gene and therefore chose the DU151 *pol* gene for further development. Only one of the recent seroconverter sequences was less than 93% identical with the DU151 *pol* gene segment.

The *env* gene showed the greatest sequence diversity. Based on the analysis of the phylogenetic tree shown in Figure 10, we chose the DU151 *env* gene. The DU151 *env* gene segment shows an average pairwise comparison score with the other isolates of 87.2%, with the clone being slightly higher (87.9%). The DU151 isolate gene segment has a pairwise identity score of 92.6% with the SA consensus while the DU151 clone is at 91.3%. Finally, all pairwise identity scores are above 83% with either the DU151 isolate sequence or the clone when compared to the other recent seroconverters, as shown in Figure 13. These pairwise scores make the DU151 sequence similar to the best scores in this sequence pool and combine these levels of similarity with an R5 virus with good cell culture replication kinetics.

The clones representing the full length gene for each of the above viral genes were generated by PCR. Viral DNA present in cells infected with the individual isolates were used for the *pol* and *env* clones, and DNA derived directly from plasma by RT-PCR was used for the *gag* clone. Total DNA was extracted from the infected cell pellets using the QIAGEN DNeasy Tissue Kit. This DNA was used in PCR reactions using the following primers (HXBR numbering, Los Alamos database) in a nested PCR amplification strategy:

gag: outer,623-640, and 2391-2408. inner, 789-810 and 2330-2350
pol: outer,2050 -2073, and 5119-5148. inner,2085 -2108, and 5068-5094.
env: outer, 6195-6218, and 8807-8830. inner, 6225-6245, and 8758-8795.

The PCR products were blunt-end cloned into pT7Blue using the Novagen pT7Blue Blunt Kit. The inserts were characterized by doing colony PCR to identify clones with gene inserts. The identity of the insert was confirmed by sequencing the insert on both strands and comparing this sequence to the original sequence.

Modification of clones

Several modifications were introduced to the cloned genes, as shown in Figures 23 to 28. In order to increase levels of expression of proteins, the DNA sequence was resynthesized and the following modifications were made:

- the codon usage was changed to reflect human codon usage for increased expression; and
- the inhibitory and rev responsive elements were also removed.

The modifications to the *gag* gene sequence of Du422 are shown in Sequence I.D. numbers 7 and 8.

Also for the DNA, modified vaccinia ankara (MVA) and BCG vaccines, the *pol* gene was truncated so that only the protease, reverse transcriptase and RNAse H regions of the *pol* gene will be expressed. In addition, the active site amino acid motive YMDD has been mutated to YMAA so that the expressed reverse transcriptase will be catalytically inactive. The modifications to the *pol* gene of Du151 are shown in sequence I.D. numbers 9 and 10.

Synthetic genes

The complete *gag* and *env* genes were resynthesized to optimise the codons for expression in human cells, also shown in Sequence I.D. numbers 9 to 12. During this process the inhibitory sequences (INS) and rev responsive elements (RRE) are removed which has reported to result in increased expression. The *gag* gene myristylation signal was mutated as described above and as shown in Sequence I.D. numbers 7 and 8.

The following material has been deposited with the European Collection of Cell Cultures, Centre for Applied Microbiology and Research, Salisbury, Wiltshire SP4 OJG, United Kingdom (ECACC).

Deposits

Material	ECACC Deposit No.	Deposit Date
HIV-1 Viral isolate Du151	Accession Number 00072724	27 July 2000
HIV-1 Viral isolate Du179	Accession Number 00072725	27 July 2000
HIV-1 Viral isolate Du422	Provisional Accession Number 00072726	27 July 2000
	Provisional Accession Number 01032114	22 March 2001

The deposit was made under the provisions of the Budapest Treaty on the International Recognition of the Deposit of Microorganisms for the Purpose of Patent Procedure and regulations thereunder (Budapest Treaty).

REFERENCES

UNAIDS. AIDS epidemic update. December 1999. www.unaids.org/hivaidsinfo/documents.html

Binley JM, Sanders RW, Clas B, Schuelke N, Master A, Guo Y, Kajumo F, Anselma DJ, Maddon PJ, Olson WC, Moore JP., J Virol 2000 Jan;74(2):627-43

Bjorndal, A., Sonnerborg, A., Tscherning, C., Albert, J. & Fenyo, E. M. (1999). Phenotypic characteristics of human immunodeficiency virus type 1 subtype C isolates of Ethiopian.

Connor, R., Sheridan, K., Ceraldini, D., Choe, S. & Landau, N. (1997). Changes in co-receptor use correlates with disease progression in HIV-1-infected individuals. *J Exp Med* **185**, 621-628.

Durali D, Morvan J, Letourneur F, Schmitt D, Guegan N, Dalod M, Saragosti S, Sicard D, Levy JP & Gomard E (1998). Cross-reactions between the cytotoxic T-lymphocyte responses of human immunodeficiency virus-infected African and European patients. *J Virol* 72:3547-53.

Ferrari G, Humphrey W, McElrath MJ, Excler JL, Duliege AM, Clements ML, Corey LC, Bolognesi DP & Weinhold KJ (1997). Clade B-based HIV-1 vaccines elicit cross-clade cytotoxic T lymphocyte reactivities in uninfected volunteers. *Proc Natl Acad Sci U S A* 18;94(4):1396-401.

HIV Molecular Immunology Database 1998: Korber B, Brander C, Koup R, Walker B, Haynes B, & Moore J, Eds. Theoretical Biology and Biophysics Group, Los Alamos National Laboratory, Los Alamos, NM.

Kostrikis, L.G., Cao, Y., Ngai, H., Moore, J.P. & Ho, D.D (1996). Quantitative analysis of serum neutralization of human immunodeficiency virus type 1 from subtypes A, B, C, D, E, F, and I: lack of direct correlation between neutralization serotypes and genetic

subtypes and evidence for prevalent serum-dependent infectivity enhancement. *J. Virol.* **70**, 445–458.

Koup RA, Safrit JT, Cao Y, Andrews CA, McLeod G, Borkowsky W, Farthing C, Ho DD (1994). Temporal association of cellular immune responses with the initial control of viremia in primary human immunodeficiency virus type 1 syndrome. *J Virol.* 68(7):4650-5.

Moore JP, Cao Y, Leu J, Qin L, Korber B & Ho DD (1996). Inter- and intraclade neutralization of human imunodeficiency virus type 1: genetic clades do not correspond to neutralization serotypes but partially correspond to gp120 antigenic serotypes. *J. Virol.* **70**, 427–444.

Ogg GS, Kostense S, Klein MR, Jurriaans S, Hamann D, McMichael AJ & Miedema F (1999). Longitudinal phenotypic analysis of human immunodeficiency virus type 1-specific cytotoxic T lymphocytes: correlation with disease progression. *J Virol*; **73**(11):9153-60.

Peeters, M., Vincent, R., Perret, J.-L., Lasky, M., Patrel, D., Liegeois, F., Courgnaud, V., Seng, R., Matton, T., Molinier, S. & Delaporte, E. (1999). Evidence for differences in MT2 cell tropism according to genetic subtypes of HIV-1: syncitium-inducing variants seem rare among subtype C HIV-1 viruses. *J Acquir Imm Def Synd* **20**, 115-121.

Richman, D. & Bozzette, S. (1994). The impact of the syncytium-inducing phenotype of human immunodeficiency virus on disease progression. *J Inf Dis* **169**, 968-974.

Robertson DL, Anderson JP, Bradac JA, Carr JK, Foley B, Funkhouser RK, Gao R, Hahn BH, Kalish ML, Kuiken C, Learn GH Leitner T, McCutchan F, Osmanov S, Peeters M, Pieniazek D, Salminen M, Sharp PM, Wolinsky S, Korber B (2000). HIV nomenclature proposal. *Science* **7;288** (5463):55-6.

Rowland-Jones SL, Dong T, Fowke KR, Kimani J, Krausa P, Newell H, Blanchard T, Ariyoshi K, Oyugi J, Ngugi E, Bwayo J, MacDonald KS, McMichael AJ & Plummer FA

(1998). Cytotoxic T-cell responses to multiple conserved epitopes in HIV-resistant prostitutes in Nairobi. *J. Clin. Invest.* **102** (9): 1758–1765.

Scarlatti, G., Tresoldi, E., Bjorndal, A., Fredriksson, R., Colognesi, C., Deng, H., Malnati, M., Plebani, A., Siccardi, A., Littman, D., Fenyo, E. & Lusso, P. (1997). In vivo evolution of HIV-1 co-receptor usage and sensitivity to chemokine-mediated suppression. *Nat Med* **3**, 1259-1265.

Schmitz JE, Kuroda MJ, Santra S, Sasseville VG, Simon MA, Lifton MA, Racz P, Tenner-Racz K, Dalesandro M, Scallon BJ, Ghrayeb J, Forman MA, Montefiori DC, Rieber EP, Letvin NL, Reimann KA (1999). Control of viremia in simian immunodeficiency virus infection by CD8+ lymphocytes. *Science* **5;283**(5403):857-60.

Summary Report: National HIV sero-prevalence survey of women attending public antenatal clinics in South Africa, 1999 (2000). Department of Health, Directorate: Health Systems Research & Epidemiology, April 2000.

Tscherning, C., Alaeus, A., Fredriksson, R., Bjorndal, A., Deng, H., Littman, D., Fenyo, E. M. & Alberts, J. (1998). Differences in chemokine co-receptor usage between genetic subtypes of HIV-1. *Virology* **241**, 181-188.

Wyatt R and Sodroski J (1998). The HIV-1 *envelope* glycoproteins: Fusogens, antigens and immunogens. *Science*, 280 (5371):1884-8.

Wyatt R, Kwong, Desjardins E, Sweet RW, Robinson J, Hendrickson WA & Sodroski JG (1998). The antigenic structure of the HIV gp120 *env*elope glycoprotein. *Nature*, 393(6686):705-11.

0-1	Form - PCT/RO/134 (EASY) Indications Relating to Deposited Microorganism(s) or Other Biological Material (PCT Rule 13bis)	
0-1-1	Prepared using	PCT-EASY Version 2.91 (updated 01.03.2001)
0-2	International Application No.	
0-3	Applicant's or agent's file reference	PA128340/PCT

1	The indications made below relate to the deposited microorganism(s) or other biological material referred to in the description on:	
1-1	page	3
1-2	line	17
1-3	Identification of Deposit	
1-3-1	Name of depositary institution	European Collection of Cell Cultures
1-3-2	Address of depositary institution	Vaccine Research and Production Laboratory, Public Health Laboratory Service, Centre for Applied Microbiology and Research, Porton Down, Salisbury, Wiltshire SP4 0JG, United Kingdom
1-3-3	Date of deposit	27 July 2000 (27.07.2000)
1-3-4	Accession Number	ECACC 00072724
1-4	Additional Indications	NONE
1-5	Designated States for Which Indications are Made	all designated States
1-6	Separate Furnishing of Indications These indications will be submitted to the International Bureau later	NONE
2	The indications made below relate to the deposited microorganism(s) or other biological material referred to in the description on:	
2-1	page	3
2-2	line	21
2-3	Identification of Deposit	
2-3-1	Name of depositary institution	European Collection of Cell Cultures
2-3-2	Address of depositary institution	Vaccine Research and Production Laboratory, Public Health Laboratory Service, Centre for Applied Microbiology and Research, Porton Down, Salisbury, Wiltshire SP4 0JG, United Kingdom
2-3-3	Date of deposit	27 July 2000 (27.07.2000)
2-3-4	Accession Number	ECACC 0072725
2-4	Additional Indications	NONE
2-5	Designated States for Which Indications are Made	all designated States

2-6	Separate Furnishing of Indications These indications will be submitted to the International Bureau later	NONE
3	The indications made below relate to the deposited microorganism(s) or other biological material referred to in the description on:	
3-1	page	3
3-2	line	13
3-3	Identification of Deposit	
3-3-1	Name of depositary institution	European Collection of Cell Cultures
3-3-2	Address of depositary institution	Vaccine Research and Production Laboratory, Public Health Laboratory Service, Centre for Applied Microbiology and Research, Porton Down, Salisbury, Wiltshire SP4 0JG, United Kingdom
3-3-3	Date of deposit	22 March 2001 (22.03.2001)
3-3-4	Accession Number	ECACC 01032114
3-4	Additional Indications	NONE
3-5	Designated States for Which Indications are Made	all designated States
3-6	Separate Furnishing of Indications These indications will be submitted to the International Bureau later	NONE
4	The indications made below relate to the deposited microorganism(s) or other biological material referred to in the description on:	
4-1	page	26
4-2	line	11
4-3	Identification of Deposit	
4-3-1	Name of depositary institution	European Collection of Cell Cultures
4-3-2	Address of depositary institution	Vaccine Research and Production Laboratory, Public Health Laboratory Service, Centre for Applied Microbiology and Research, Porton Down, Salisbury, Wiltshire SP4 0JG, United Kingdom
4-3-3	Date of deposit	27 July 2000 (27.07.2000)
4-3-4	Accession Number	ECACC 00072726
4-4	Additional Indications	NONE
4-5	Designated States for Which Indications are Made	all designated States
4-6	Separate Furnishing of Indications These indications will be submitted to the International Bureau later	NONE

Original (for SUBMISSION) - printed on 06.07.2001 04:52:24 PM

FOR RECEIVING OFFICE USE ONLY

0-4	This form was received with the international application: (yes or no)	yes
0-4-1	Authorized officer	M. Cavalleri

FOR INTERNATIONAL BUREAU USE ONLY

0-5	This form was received by the International Bureau on:	
0-5-1	Authorized officer	

CLAIMS

1. A process for the selection of HIV subtype isolates for use in the development of a prophylactic and/or therapeutic pharmaceutical composition comprising the following steps:
isolating viruses from recently infected subjects;
generating a consensus sequence for at least part of at least one HIV gene by identifying the most common codon or amino acid among the isolated viruses at each position along at least part of the gene;
selecting the isolated virus or viruses with a high sequence identity to the consensus sequence, a phenotype which is associated with transmission for the particular HIV subtype.

2. A process according to claim 1, wherein the isolated virus is of the same subtype as a likely challenge strain.

3. A process according to either of claims 1 or 2, wherein the HIV subtype is HIV-1 subtype C.

4. A process according to claim 3, wherein the phenotype which is associated with transmission is a virus that utilises the CCR5 co-receptor and is non syncitium inducing (NSI).

5. An HIV-1 subtype C isolate, designated Du422 and assigned Provisional Accession Number 01032114 by the European Collection of Cell Cultures.

6. An HIV-1 subtype C isolate, designated Du151 and assigned Accession Number. 00072724 by the European Collection of Cell Cultures.

7. An HIV-1 subtype C isolate, designated Du179 and assigned Accession Number. 00072725 by the European Collection of Cell Cultures.

8. A molecule having:
(i) the nucleotide sequence set out in Sequence I.D. No 1;

(ii) an RNA sequence corresponding to the nucleotide sequence set out in Sequence I.D. No 1;

(iii) a sequence which will.hybridise to the nucleotide sequence set out in Sequence I.D. No 1 or an RNA sequence corresponding to it, under strict hybridisation conditions;

(iv) a sequence which is homologous to the nucleotide sequence set out in Sequence I.D. No 1 or an RNA sequence corresponding to it; or

(v) a sequence which is a modification or derivative of the sequence of any one of (i) to (iv).

9. A molecule according to claim 8, which has the modified sequence set out in Sequence I.D. No 7.

10. A molecule having:

(i) the nucleotide sequence set out in Sequence I.D. No 3;

(ii) an RNA sequence corresponding to the nucleotide sequence set out in Sequence I.D. No 3;

(iii) a sequence which will hybridise to the nucleotide sequence set out in Sequence I.D. No 3 or an RNA sequence corresponding to it, under strict hybridisation conditions;

(iv) a sequence which is homologous to the nucleotide sequence set out in Sequence I.D. No. 3 or an RNA sequence corresponding to it; or

(v) a sequence which is a modification or derivative of the sequence of any one of (i) to (iv).

11. A molecule according to claim 10, which has the modified sequence set out in Sequence I.D. No. 9.

12. A molecule having:

(i) the nucleotide sequence set out in Sequence I.D. No. 5;

(ii) an RNA sequence corresponding to the nucleotide sequence set out in Sequence I.D. No. 5;

(iii) a sequence which will hybridise to the nucleotide sequence set out in Sequence I.D. No. 5 or an RNA sequence corresponding to it, under strict hybridisation conditions;

(iv) a sequence which is homologous to the nucleotide sequence set out in Sequence I.D. No. 5 or an RNA sequence corresponding to it; or

(v) a sequence which is a modification or derivative of the sequence of any one of (i) to (iv).

13. A molecule according to claim 12, which has the modified sequence set out in Sequence I.D. No. 11.

14. A molecule having:

(i) the nucleotide sequence set out in Sequence I.D. No.13;

(ii) an RNA sequence corresponding to the nucleotide sequence set out in Sequence I.D. No. 13;

(iii) a sequence which will hybridise to the nucleotide sequence set out in Sequence I.D. No. 13 or an RNA sequence corresponding to it, under strict hybridisation conditions;

(iv) a sequence which is homologous to the nucleotide sequence set out in Sequence I.D. No. 13 or an RNA sequence corresponding to it; or

(v) a sequence which is a modification or derivative of the sequence of any one of (i) to (iv).

15. A molecule according to claim 14, which has a modified sequence which has similar or the same modifications as those set out in Sequence I.D. No. 11 for the *env* gene of the isolate Du151.

16. A polypeptide having:

(i) the amino acid sequence set out in Sequence I.D. No. 2; or

(ii) a sequence which is a modification or derivative of the amino acid sequence set out in Sequence I.D. No. 2.

17. A polypeptide according to claim 16, wherein the modified sequence is set out in Sequence I.D. No. 8.

18. A polypeptide having:
    (i) the amino acid sequence set out in Sequence I.D. No. 4; or
    (ii) a sequence which is a modification or derivative of the amino acid sequence set out in Sequence I.D. No. 4.

19. A polypeptide according to claim 18, wherein the modified sequence is that set out in Sequence I.D. No. 10.

20. A polypeptide having:
    (i) the amino acid sequence set out in Sequence I.D. No. 6; or
    (ii) a sequence which is a modification or derivative of the amino acid sequence set out in Sequence I.D. No. 6.

21. A polypeptide according to claim 20, wherein the modified sequence is that set out in Sequence I.D. No. 12.

22. A polypeptide having:
    (i) the amino acid sequence set out in Sequence I.D. No. 14;
    (ii) a sequence which is a modification or derivative of the amino acid sequence set out in Sequence I.D. No. 14.

23. A polypeptide according to claim 22, wherein the modified sequence has similar or the same modifications as those set out in Sequence I.D. No. 12 for the amino acid sequence of the *env* gene of the isolate Du151.

24. A consensus amino acid sequence for the partial *gag* gene of HIV-1 subtype C which is:

GEKLDKWEKI	RLRPGGKKHY	MLKHLVWASR	ELERFALNPG	LLETSEGCKQ [50]
IMKQLQPALQ	TGTEELRSLY	NTVATLYCVH	EKIEVRDTKE	ALDKIEEEQN [100]
KSQQ-CQQKT	QQAKAADGG-	KVSQNYPIVQ	NLQGQMVHQA	ISPRTLNAWV [150]
EEKAFSP	EVIPMFTALS	EGATPQDLNT	MLNTVGGHQA	AMQMLKDTIN [200]
EEAAEWDRLH	PVHAGPIAPG	QMREPRGSDI	AGTTSTLQEQ	IAWMTSNPPI [250]

PVGDIYKRWI ILGLNKIVRM YSPVSILDIK QGPKEPFRDY VDRFFKTLRA [300]
EQATQDVKNW MTD [313]

25. A consensus amino acid sequence for the partial *pol* gene of HIV-1 subtype C which is:

LTEEKIKALT AICEEMEKEG KITKIGPENP YNTPVFAIKK KDSTKWRKL- [50]
VDFRELNKRT QDFWEVQLGI PHPAGLKKKK SVTVLDVGDA YFSVPLDEGF [100]
RKYTAFTIPS INNETPGIRY QYNVLPQGWK GSPAIFQSSM TKILEPFRAK [150]
NPEIVIYQYM DDLYVGSDLE IGQHRAKIEE LREHLLKWGF TTPDKKHQKE [200]
PPFLWMGYEL HPDKWTVQPI QLPEKDSWTV NDIQKLVGKL NWASQIYPGI [250]
KVRQLCKLLR GAKALTDIVP LTEEAELE [278]

26. A consensus amino acid sequence for the partial *env* gene of HIV-1 subtype C which is:

YCAPAGYAIL KCNNKTFNGT GPCNNVSTVQ CTHGIKPVVS TQLLLNGSLA [50]
EEEIIIRSEN LTNNAKTIIV HLNESVEIVC TRPNNNTRKS IRIGPGQTFY [100]
ATGDIIGDIR QAHCNISEGK WNKTLQKVKK KLKEELYKYK VVEIKPLGIA [150]
PTEAKRRVVE REKRAVGIGA VFLGFLGAAG STMGAASITL TVQARQLLSG [200]
IVQQQSNLLR AIEAQQHMLQ LTVWGIKQL [229]

27. A process according to claim 1, substantantially as herein described.

28. An HIV-1 subtype C isolate according to claim 5, substantially as herein described.

29. An HIV-1 subtype C isolate according to claim 6, substantially as herein described.

30. An HIV-1 subtype C isolate according to claim 7, substantially as herein described.

31. A molecule according to claim 8, substantially as herein described.

32. A molecule according to claim 10, substantially as herein described.

33. A molecule according to claim 12, substantially as herein described.

34. A molecule according to claim 14, substantially as herein described.

35. A polypeptide according to claim 16, substantially as herein described.

36. A polypeptide according to claim 18, substantially as herein described.

37. A polypeptide according to claim 20, substantially as herein described.

38. A polypeptide according to claim 22, substantially as herein described.

39. A consensus amino acid sequence according to claim 24, substantially as herein described.

40. A consensus amino acid sequence according to claim 25, substantially as herein described.

41. A consensus amino acid sequence according to claim 26, substantially as herein described.

FIGURE 1

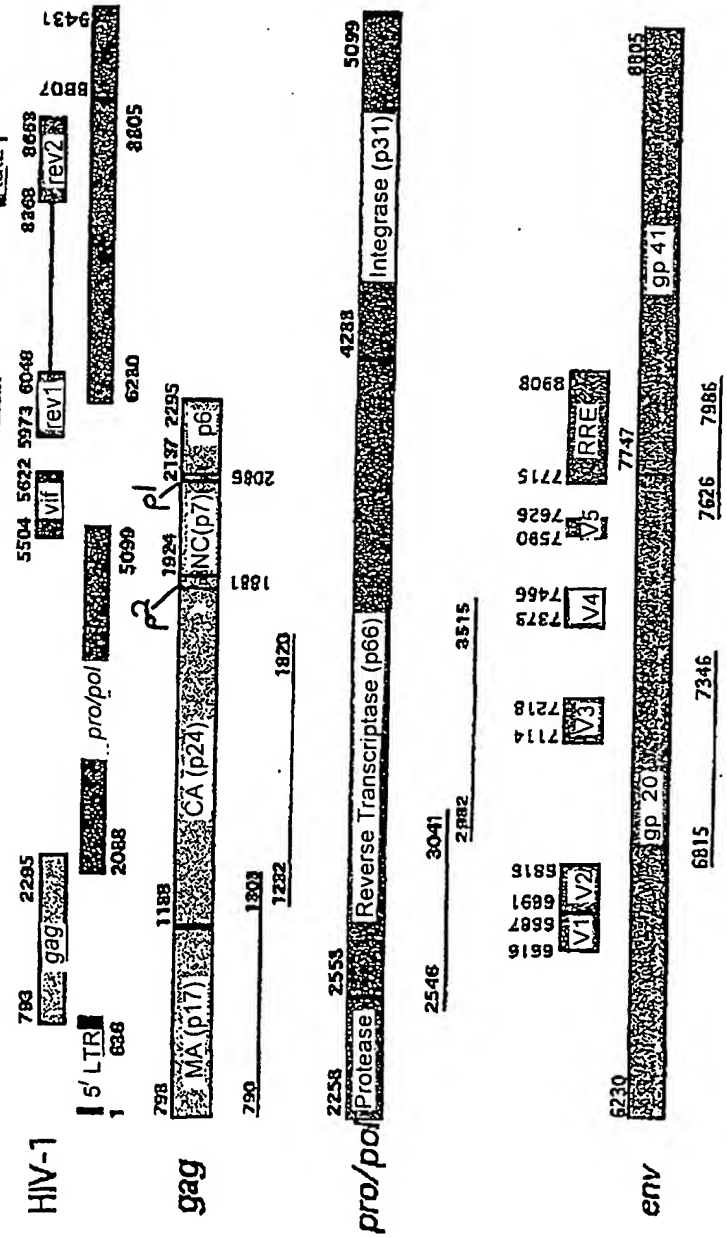

HIV-1
5' LTR
gag
pro/pol
MA (p17)
CA (p24)
NC(p7)
p6
Protease
Reverse Transcriptase (p66)
Integrase (p31)
gp 120
gp 41
RRE
vpr
vpu
tat1
tat2
rev1
rev2
vif
V1
V2
V3
V4
V5

FIGURE 2

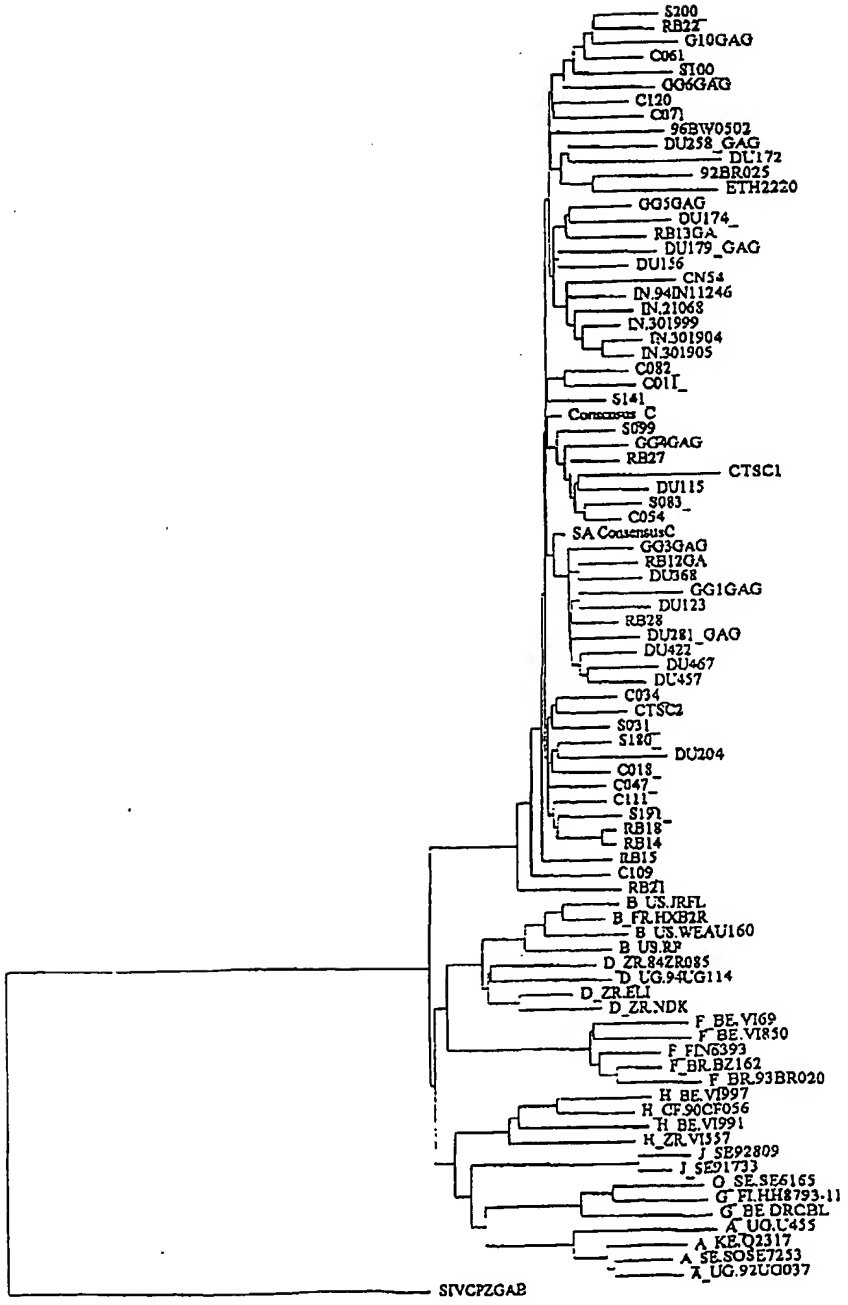

S200
RB22
G10GAG
C061
SI00
GG6GAG
C120
C071
96BW0502
DU258_GAG
DU172
92BR025
ETH2220
GG5GAG
DU174_
RB13GA
DU179_GAG
DU156
CN54
IN.94IN11246
IN.21068
IN.301999
IN.301904
IN.301905
C082
C011_
S141
Consensus C
S099
GG2GAG
RB27
CTSC1
DU115
S083_
C054
SA ConsensusC
GG3GAG
RB12GA
DU368
GG1GAG
DU123
RB28
DU281_GAG
DU422
DU467
DU457
C034
CTSC2
S031
S180_
DU204
C018_
C047_
C111_
S191_
RB18
RB14
RB15
C109
RB21
B_US.JRFL
B_FR.HXB2R
B_US.WEAU160
B_US.RF
D_ZR.84ZR085
D_UG.94UG114
D_ZR.ELI
D_ZR.NDK
F_BE.VI69
F_BE.VI850
F_FIN6393
F_BR.BZ162
F_BR.93BR020
H_BE.VI997
H_CF.90CF056
H_BE.VI991
H_ZR.VI557
J_SE92809
J_SE91733
O_SE.SE6165
G_FI.HH8793-11
G_BE.DRCBL
A_UG.U455
A_KE.Q2317
A_SE.SOSE7253
A_UG.92UG037
SIVCPZGAB

FIGURE 3

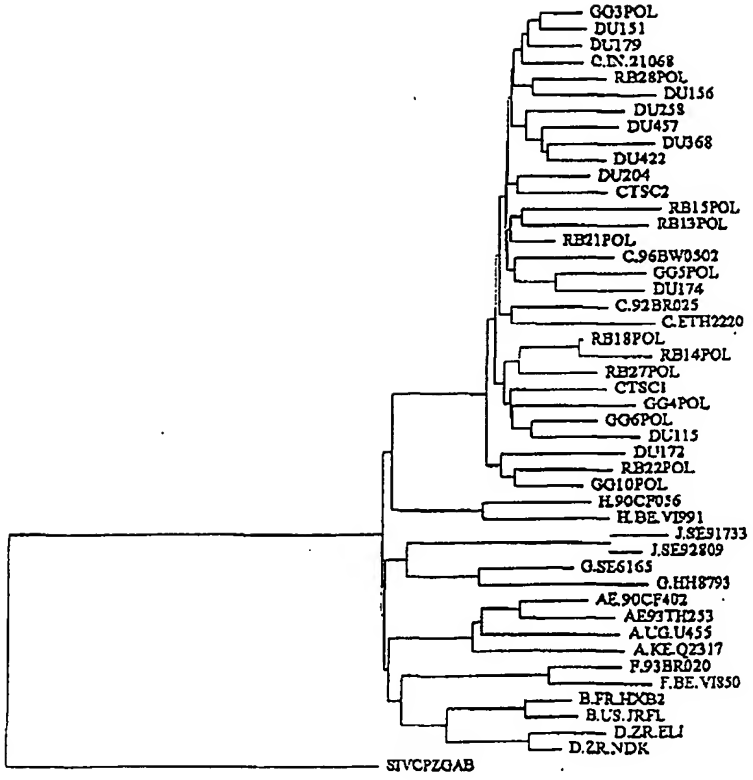

GG3POL
DU151
DU179
C.IN.21068
RB28POL
DU156
DU258
DU457
DU368
DU422
DU204
CTSC2
RB15POL
RB13POL
RB21POL
C.96BW0502
GG5POL
DU174
C.92BR025
C.ETH2220
RB18POL
RB14POL
RB27POL
CTSC1
GG4POL
GG6POL
DU115
DU172
RB22POL
GG10POL
H.90CF056
H.BE.VI991
J.SE91733
J.SE92809
G.SE6165
G.HH8793
AE.90CF402
AE93TH253
A.UG.U455
A.KE.QZ317
F.93BR020
F.BE.VI850
B.FR.HXB2
B.US.JRFL
D.ZR.ELI
D.ZR.NDK
SIVCPZGAB

FIGURE 4

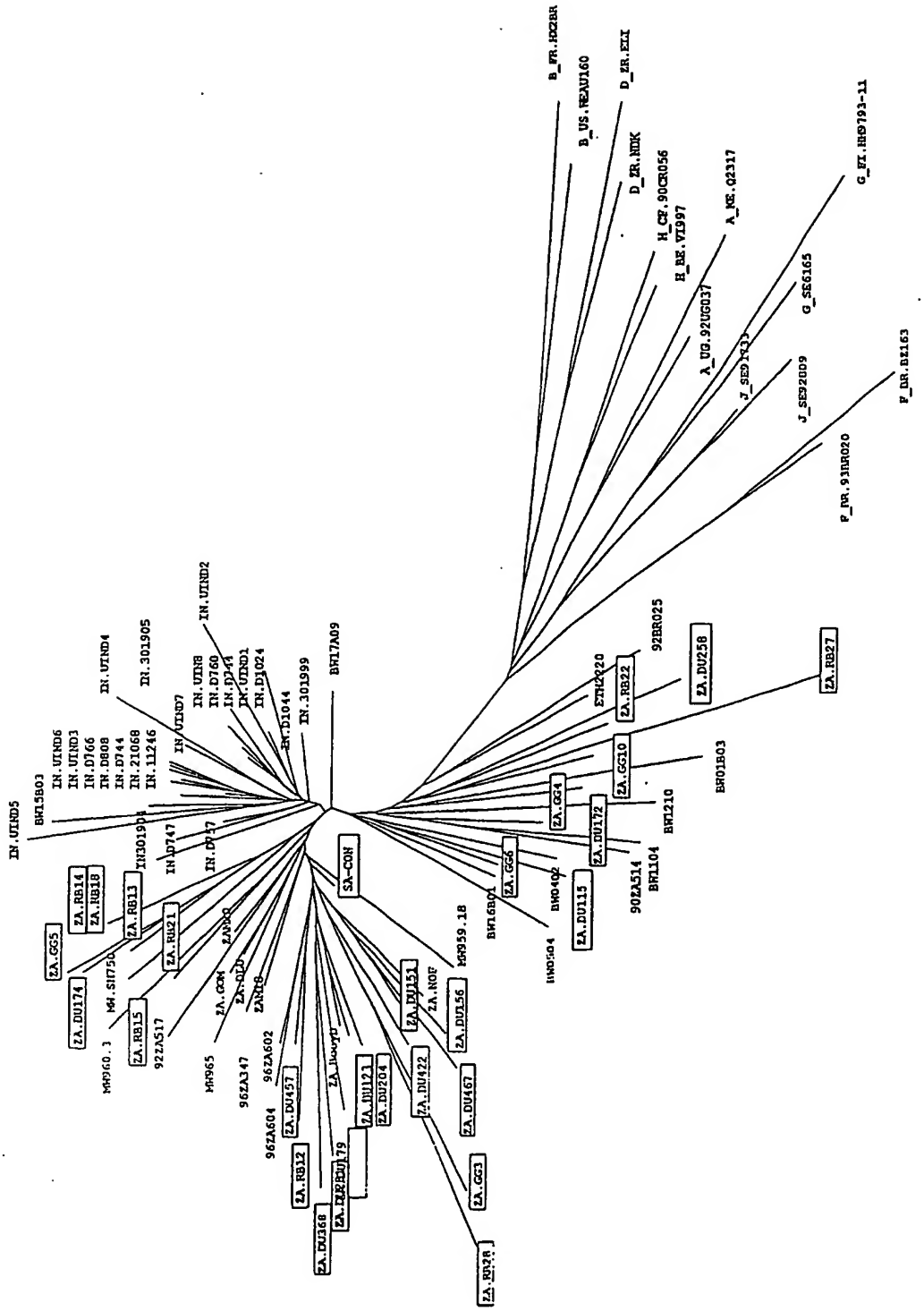

B_FR.HXB2R
B_US.NEAU160
D_ZR.EXI
D_ZR.NDK
H_CF.90CR056
H_BE.VI997
A_UG.92UG037
A_KE.02317
J_SE9173-11
G_SE6165
J_SE92809
F_FR.93FR020
F_BR.BZ163
IN.UIND5
BH15B03
IN.UIND6
IN.UIND3
IN.D766
IN.D808
IN.D744
IN.21068
IN.11246
IN.UIND4
IN.301905
IN.UIND7
IN.UIND8
IN.D760
IN.D744
IN.UIND2
IN.UIND1
IN.D1024
IN.D1044
IN.301999
BH17A09
IN301909
IN.D747
IN.D737
ZA.GG5
ZA.RB14
ZA.RB18
ZA.RB13
ZA.RB21
ZA.DU174
M4.SIM750
ZA.RB15
922A517
M4960.3
M4965
ZA.CON
ZA.CLD
ZA.DU
ZARD8
962A347
962A602
962A604
ZA.DU457
ZA.RB12
ZA.DU368
ZA.DURBDU179
ZA.DU121
ZA.DU204
ZA.DU422
ZA.DU151
ZA.NOF
ZA.DU156
ZA.DU467
ZA.GG3
ZA.RB28
SA-CON
M4959.18
BH16A01
BH0402
BH0504
ZA.GG6
ZA.DU115
902A514
BH1104
ZA.DU172
ZA.GG4
BH1210
ZA.GG10
BH01B03
ZTH2220
ZA.RB22
92BR025
ZA.DU258
ZA.RB27

FIGURE 5


```
                          1                                                  50
 @Gagp.msf{SAgagcon}  GEKLDKWEKI RLRPGGKKHT MLKHLVWASR ELERFALNPG LLETSEGCKQ
@Gagp.msf{DU115_gag}  -q---d--t- ---------t- -i--------- ---------- ----g-----
@Gagp.msf{DU258_gag}  -q-------- k--------- ---------- ---------- ----ad----
@Gagp.msf{DU179_gag}  -q-------- ---------- -i--------- ---------- --------r-
  @Gagp.msf{RB15gag}  -g---a---- ---------- -i--------- ---------- -----a----
  @Gagp.msf{GG10gag}  ---------- ---------- ---------- ---------- ----------
  @Gagp.msf{RB22gag}  ---------- ---------q- ---------- ---------- ----------
@Gagp.msf{DU467_gag}  -----k---- ---------r- ----i----- ---------- ----------
   @Gagp.msf{GG3gag}  -----d---- ---------- ----i----- ---------- ----------
@Gagp.msf{DU281_gag}  --------r- ---------e- ----i----- ---------- ----------
@Gagp.msf{DU368_gag}  -----r--r- ---------- ----i----- ---------- --------r-
   @Gagp.msf{GG1gag}  -----r---- ---------- ----i----- ---------- ----------
  @Gagp.msf{RB14gag}  -g---a---- --------c- -i---i---- ---------- ----------
  @Gagp.msf{RB18gag}  -----a---- k-------c- -i--------- ---------- ----------
   @Gagp.msf{GG4gag}  -g---a---- ---------- -i--------- ---------- ----------
  @Gagp.msf{RB27gag}  -gq------- ---------- ---------- ---------- ----------
@Gagp.msf{DU422_gag}  -----t---- ---------- ----i----- ---------- ----------
  @Gagp.msf{RB28gag}  -----t--r- ---------- ----i----- ---------- ----------
@Gagp.msf{DU457_gag}  --n--r---- --------r- ----i----- ---------- ----------
  @Gagp.msf{RB13gag}  -k---s--r- ---------- -i--------- ---------- ----------
   @Gagp.msf{GG5gag}  -g---t---- --------q- -i--------- ---------- --------r-
   @Gagp.msf{GG6gag}  ---------- --------k- -----i---- ---------- ----a-----
@Gagp.msf{CTSC2_gag}  -g-------- --------c- ----ii---- ---------- ----k-----
  @Gagp.msf{RB12gag}  -----r--r- --------t- ----i----- ---------- ----------
    @Gagp.msf{DU156}  ---------- ---------- ---------- ---------d ----------
@Gagp.msf{DU123_gag}  --------r- k--------- r---i----- ---------- ----------
  @Gagp.msf{RB21gag}  -g---t---- --------r- km--i----- ---------- ----a-----
@Gagp.msf{DU172_gag}  -g------r- ---------- -----i---- ---------- ----------
@Gagp.msf{DU204_gag}  -g-------- --------r- -i--------- ---------- ----a-----
@Gagp.msf{DU174_gag}  -gn--t---- --------q- k--------- ---------- ---sa-----
@Gagp.msf{CTSC1_gag}  -g---a--r- ---------- -----i---k ---------- ---n------
  @Gagp.msf{Cgagcon}  -g---t---- ---------- -i--------- ---------- ----------

                         51                                                 100
 @Gagp.msf{SAgagcon}  IMKQLQPALQ TGTEELRSLY NTVATLYCVH EKIEVRDTKE ALDKIEEEQN
@Gagp.msf{DU115_gag}  ---------- ---k------ ---------- ---------- ---------k
@Gagp.msf{DU258_gag}  -i-------- ---------f ---------- ke-------- ----------
@Gagp.msf{DU179_gag}  -ir------- ---------f ---------- -e-------- ---r-----k
  @Gagp.msf{RB15gag}  -ir------- ---------- ---------- -r-------- ---------k
  @Gagp.msf{GG10gag}  --s---s--- ---------- ----v-w--- nn-------- ---------k
  @Gagp.msf{RB22gag}  ---------- ---------- ---------- sn-------- ----------
@Gagp.msf{DU467_gag}  --e------- ------k--f ---------- kr-d------ ----v----k
   @Gagp.msf{GG3gag}  --n------- ---------- ---------- kr-d------ ----------
@Gagp.msf{DU281_gag}  --q------- ------k--- --i------- kg-------- ----------
@Gagp.msf{DU368_gag}  --n------- ------k--- ---------- ---d------ ----------
   @Gagp.msf{GG1gag}  -it------- ------k--f ---------- k--d------ ----------
  @Gagp.msf{RB14gag}  ---------- ---------- ---------- ---------- ----------
  @Gagp.msf{RB18gag}  ---------- ---------- ---------- -----q---- ----------
   @Gagp.msf{GG4gag}  -i-------- ---------- ---------- ---d------ ----------
  @Gagp.msf{RB27gag}  -i-------- ---------- ---------- ---------- ----------
@Gagp.msf{DU422_gag}  ---------- ------k--- ---------- ---------- ----------
  @Gagp.msf{RB28gag}  -ir------- ------k--f ---------- ---k------ ----------
@Gagp.msf{DU457_gag}  ---------- ------k--- ---------- k--d-q---- ----v-----
  @Gagp.msf{RB13gag}  ---------- ---------f ---------- ---------- ----------
   @Gagp.msf{GG5gag}  ---------- ---------f ---------- ---------- ----------
   @Gagp.msf{GG6gag}  --q-i----- ---------f ---------- .--------- ----------
@Gagp.msf{CTSC2_gag}  -in--h---- ---------f ---------- de-------- ----------
  @Gagp.msf{RB12gag}  --n-----v- ------k--f ---------- ---d------ ----------
```

FIGURE 5 - continue


```
@Gagp.msf{DU156}  ---------- ---------r ---------- ----i----- ----------
@Gagp.msf{DU123_gag}  --n------- ---------t ---------- ----d----- ----------
@Gagp.msf{RB21gag}  -iq------k ------s--t ---------- kr-------- ----v-----
@Gagp.msf{DU172_gag}  -i---h---- ---------- --i-y----- kd-a-q---- ----------
@Gagp.msf{DU204_gag}  -iq------k ------k--- ---------- ne-------- ----------
@Gagp.msf{DU174_gag}  -i-------- ---------- ---------- -r---q---- ----------
@Gagp.msf{CTSC1_gag}  -----h---- ------k--- ---------- ---------- ----------
@Gagp.msf{Cgagcon}  -i-------- ---------- ---------- ---------- ----------

                    101                                                 150
@Gagp.msf{SAgagcon}  KSQQ-CQQKT QQAKAADGG- KTSQVTPIVQ NLQGQMVHQA ISPRTLNAWV
@Gagp.msf{DU115_gag}  -....----i ---e---k-. ---------- ---------- ----------
@Gagp.msf{DU258_gag}  -....----a ---e-s-k-. ---------- ---------p l---------
@Gagp.msf{DU179_gag}  --....---- ----e--k.. ---------- --------x- l---------
@Gagp.msf{RB15gag}  -c....---- ---..--e.. ---------- -a-------- ----------
@Gagp.msf{GG10gag}  --....---- --------.. n--------- ----a----- l---------
@Gagp.msf{RB22gag}  --....---- ---g----.. ---------- ---------p l---------
@Gagp.msf{DU467_gag}  r-....---- -------e.. ---------- ---------- ----------
@Gagp.msf{GG3gag}  --....---- -------ekv ---------- ---------p ----------
@Gagp.msf{DU281_gag}  --....---- --t----.-. ---------- ---------- ----------
@Gagp.msf{DU368_gag}  --....---- --e---g.-. ---------- ---------- ----------
@Gagp.msf{GG1gag}  ----.v---- ----t--.-. ---------- ---------- ----------
@Gagp.msf{RB14gag}  --....---- ---q---.-. ---------- -v-------- l---------
@Gagp.msf{RB18gag}  --....---- ---q---.-. ---------- -v-------- l---------
@Gagp.msf{GG4gag}  -i....---- ---e---k-. ---------- ---------- ----------
@Gagp.msf{RB27gag}  --....---- ---e---k-. ---------- ---------- l---------
@Gagp.msf{DU422_gag}  -c....---- -------.-. ---------- ---------- ----------
@Gagp.msf{RB28gag}  --....---- -------.e. -i-------- ---------- ----------
@Gagp.msf{DU457_gag}  --....---- ---e---.-. ---------- ---------- ----------
@Gagp.msf{RB13gag}  --....---- ----e--.-. -t-------- ---------- l---------
@Gagp.msf{GG5gag}  --....---- ----g--.-. -i-------- ---------- ----------
@Gagp.msf{GG6gag}  --....---- ------.--. q--------- ---------- ----------
@Gagp.msf{CTSC2_gag}  ...n.i---- --------.. ---------- ---------- ----------
@Gagp.msf{RB12gag}  --....---- p-------.. ---------- ---------- ----------
@Gagp.msf{DU156}  -...------ ----e---.. ---------- ---------- ----------
@Gagp.msf{DU123_gag}  --....---- -......... .--------- ---------p -t--------
@Gagp.msf{RB21gag}  --....---- ---d---k-. -----t---- -v-------- ----------
@Gagp.msf{DU172_gag}  -c--ks---- --ta----ja ..-------- ---------s l---------
@Gagp.msf{DU204_gag}  -....----- k-t-ed--.. -a-------- -a--a----p ----------
@Gagp.msf{DU174_gag}  -....i---- -----keadg -t-------- -i-------- l---------
@Gagp.msf{CTSC1_gag}  -....--h-a ---etd-k-. ---------- ---------- l---------
@Gagp.msf{Cgagcon}  --....---- -------.-. ---------- ---------- ----------

                    151                                                 200
@Gagp.msf{SAgagcon}  KVIEEKAFSP EVIPMFTALS EGATPQDLNT MLNTVGGHQA AMQMLKDTIN
@Gagp.msf{DU115_gag}  ---------- ---------- ---------- ---------- ----------
@Gagp.msf{DU258_gag}  ---------- ---------- ---------- ---------- ----------
@Gagp.msf{DU179_gag}  ---------- ---------- ---------- ---------- ----------
@Gagp.msf{RB15gag}  ---------- ---------- ---------- ---------- ----------
@Gagp.msf{GG10gag}  --------n- -i-------- ---------- ---------- ----------
@Gagp.msf{RB22gag}  --------n- ---------- ---------- ---------- ----------
@Gagp.msf{DU467_gag}  ---------- -i-------- ---------- ---------- ----------
@Gagp.msf{GG3gag}  ---------- ---------- ---------- ---------- ----------
@Gagp.msf{DU281_gag}  ---------- ---------- ---------- ---------- ----------
@Gagp.msf{DU368_gag}  ---------- ---------- ---------- ---------- ----------
@Gagp.msf{GG1gag}  ---k------ ---------- ---------- ---------- ----------
@Gagp.msf{RB14gag}  ---------- ---------- ---------- ---------- ----------
@Gagp.msf{RB18gag}  ---------- ---------- ---------- ---------- ----------
@Gagp.msf{GG4gag}  ---------- ---------- ---------- ---------- ----------
@Gagp.msf{RB27gag}  ---------- ---------- ---------- ---------- ----------
```

FIGURE 5 – continue

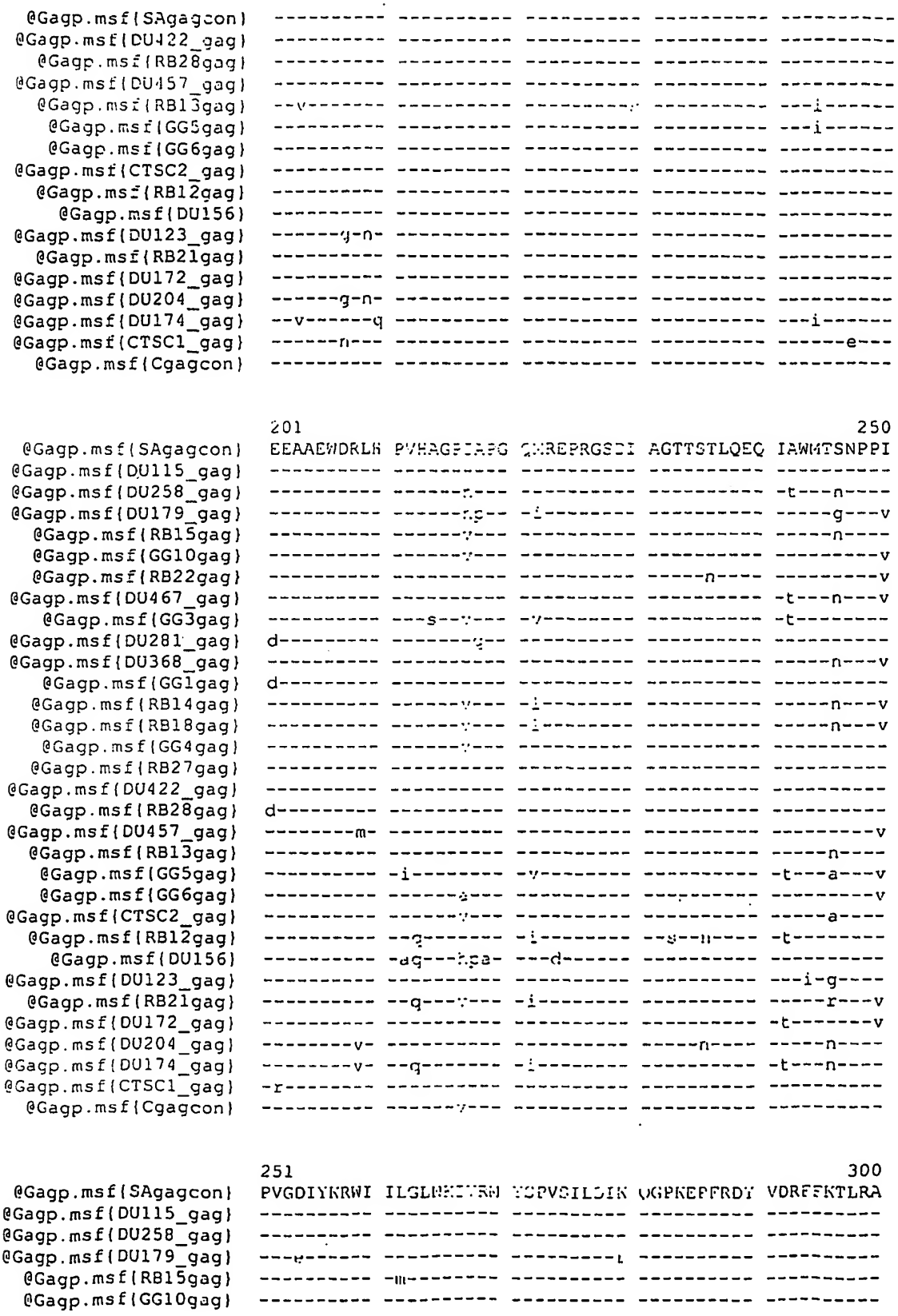


FIGURE 5 – continue

```
@Gagp.msf{RB22gag}  ---e------ ---------- ---------- ---------- ----------
@Gagp.msf{DU467_gag} ---------- ---------- ---------r ---------- ----------
@Gagp.msf{GG3gag}   ---------- ---------- ---------t ---------- ----------
@Gagp.msf{DU281_gag} ---------- ---------- ---------r ---------- ----------
@Gagp.msf{DU368_gag} ---------- ---------- ---------r ---------- ------v---
@Gagp.msf{GG1gag}   ---------- ---------- ---------t ---------- ----------
@Gagp.msf{RB14gag}  ---e------ ---------- ---------- ---------- ----------
@Gagp.msf{RB18gag}  ---e------ ---------- ---------- ---------- ----------
@Gagp.msf{GG4gag}   ---------- ---------- ---------- ---------- ----------
@Gagp.msf{RB27gag}  ---------- ---------- ---------- ---------- ----------
@Gagp.msf{DU422_gag} ---------- ---------- ---------r ---------- ----------
@Gagp.msf{RB28gag}  ---------- ---------- ---------t ---------- ----------
@Gagp.msf{DU457_gag} ---------- ---------- ---------t ---------- ----------
@Gagp.msf{RB13gag}  ---e------ ---------- ---------- ---------- ----------
@Gagp.msf{GG5gag}   ---------- ---------- ---------r ---------- ----------
@Gagp.msf{GG6gag}   ---------- ---------- ---------r ---------- ----------
@Gagp.msf{CTSC2_gag} ---------- ---------- ---------r ---------- ----------
@Gagp.msf{RB12gag}  ---------- ---------- ---------- ---------- ------a---
@Gagp.msf{DU156}    ---e------ ---------- ---------- ---------- ----------
@Gagp.msf{DU123_gag} ---e------ ---------- ---------- ---------- ----------
@Gagp.msf{RB21gag}  ---------- -m-------- ---------- ---------- ----------
@Gagp.msf{DU172_gag} ---e------ ---------- ---------- ---------- ----------
@Gagp.msf{DU204_gag} ---e------ v--------- ---------- ---------- ----------
@Gagp.msf{DU174_gag} ---e------ ---------- ---------r ---------- ----------
@Gagp.msf{CTSC1_gag} ---e------ ---------- ---------r ----ra---- ----------
@Gagp.msf{Cgagcon}  ---------- ---------- ---------- ---------- ----------

                    301        313
@Gagp.msf{SAgagcon} EQATQDVKNW MTD
@Gagp.msf{DU115_gag} --s--e---- ---
@Gagp.msf{DU258_gag} ---------- ---
@Gagp.msf{DU179_gag} ---------- ---
@Gagp.msf{RB15gag}  ---------- ---
@Gagp.msf{GG10gag}  ---------- ---
@Gagp.msf{RB22gag}  ---------- ---
@Gagp.msf{DU467_gag} -----e---- ---
@Gagp.msf{GG3gag}   ---------- --e
@Gagp.msf{DU281_gag} -----e---- ---
@Gagp.msf{DU368_gag} -----e---- --e
@Gagp.msf{GG1gag}   -----e---- ---
@Gagp.msf{RB14gag}  ---------- ---
@Gagp.msf{RB18gag}  ---------- ---
@Gagp.msf{GG4gag}   --s--e---- ---
@Gagp.msf{RB27gag}  -----e---- ---
@Gagp.msf{DU422_gag} -----e---- ---
@Gagp.msf{RB28gag}  ---------- ---
@Gagp.msf{DU457_gag} ---------- ---
@Gagp.msf{RB13gag}  -----e---- ---
@Gagp.msf{GG5gag}   ---------- ---
@Gagp.msf{GG6gag}   ---------- ---
@Gagp.msf{CTSC2_gag} ---------- --e
@Gagp.msf{RB12gag}  ---------- ---
@Gagp.msf{DU156}    ---------- ---
@Gagp.msf{DU123_gag} ---------- ---
@Gagp.msf{RB21gag}  -----e---- ---
@Gagp.msf{DU172_gag} ---------- --e
@Gagp.msf{DU204_gag} ---------- --e
@Gagp.msf{DU174_gag} ---------- ---
@Gagp.msf{CTSC1_gag} d-s--e---- ---
@Gagp.msf{Cgagcon}  ---------- ---
```

FIGURE 6

```
                    1                                                  50
sapeg.msf(SApolcon) LTEEKIKALT AICEEMEKEG KITKIGPENP YNTPVFAIKK KDSTKWRKL
ctsc1               ---------- e--------- ---------- ---------- ---------
du422               ---------- ---------- ---------- ----i----- ---------
du457               ---------- e--------- ---------- ---------- ---------
GG5pol              ---------- ----d----- ---------- ---------- ---------
ctsc2               ---------- ---------- ---------- ---------- ---------
msfdu174            -s-------- ---------- ---------- ---------- ---------
du151)              ---------- ---------- ---------- ----i----- ---------
RB27pol             ---------- ---d------ ---------- ---------- ---------
du204               ---------- e--------- ---------- ---------- ---------
RB18pol             ---------- ---------- ---------- ---------- ---------
du156               ---------- ---------- ---------- ----i----- ---------
GG3pol              -----x---- ---d------ ---------- ----i----- ---------
RB21pol             ---------- e--k------ ---------- ----i----- ---------
GG10pol             ---------k e--------- ---------- ----i----- ---------
RB28pol             -r----x--- ---------- ---------- ----i----- ---------
GG6pol              ---------- ---------- ---------- ---------- ---------
RB13pol             ---------- ---------- ---------- ---------- ---------
RB15pol             -s-------- e--------- ---------- ---------- ---------
GG4pol              ---------- e--------- ---------- -x-------- ---------
RB22pol             ---------- ---------- ---------- ---------- ---------
du172               -s-------- e--------- ---------- ---------- ---------
du115               ---------- ---------- ---------- ---------- ---------
RB14po              pavfq-sv-- ---------- ---------- ---------- ---------
du368               ---------p -f-d------ ---------- ----i----- ---------
du258               ---------- ---------- ---------- ---------- ---------v
Cpolcon             ---------- ---d------ ---------- ----i----- ---------


                    1                                                 100
SApolcon            VDFRELNKRT QDFWEVQLGI PHPAGLKKKK SVTVLDVGDA YFSVPLDEGF
ctsc1               ---------- ---------- ---------- ---------- ------h-d-
du422               ---------- ---------- ---------- ---------- ----------
du457               ---------- ---------- ---------- ---------- ----------
GG5pol              ---------- ---------- ---------- ---------- --------n-
ctsc2               ---------- ---------- ---------- ---------- --------n-
du174               ---------- ---------- ---------- ---------- --------n-
du151               ---------- ---------- ---------- ---------- ----------
RB27pol             ---------- ---------- ---------- ---------- ----------
du204               ---------- ---------- ---------- ---------- ----------
RB18pol             ---------- ---------- ---------- ---------- ----------
du156               ---------- ---------- ---------- ---------- -------pd-
du179               ---------- ---------- ---------- ---------- --------d-
GG3pol              ---------- ---------- ---------- ---------- --------n-
RB21pol             ---------- ---------- ---------- ---------- ----------
GG10pol             ---------- ---------- ---------- ---------- ----------
RB28pol             ---------- ---------- ---------- ---------- -------kd-
GG6pol              ---------- ---------- ---------r ---------- --------s-
RB13pol             ---------- ---------- ---------r ---------- --------s-
RB15pol             ---------- ---------- ---------- ---------- --------s-
GG4pol              ---------- ---------- ---------- ---------- --------s-
RB22pol             ---------- ---------- ---------- ---------- ---i--y-x-
du172               ---------- ---------- ---------- ---------- -------kd-
```

FIGURE 6 - continue

```
du115              ---------- ---------- ---------- ---------- --------n-
RB14pol            ---------- ---------- ---------- ---------- ----------
du368              ---------- ---------- ---------- ---------- ----------
du258              ---------- ---------- ---------- ---------- --------s-
Cpolcon            ---------- ---------- ---------- ---------- ------y-d-

                    101                                                150
SApolcon           RKYTAFTIPS INNETPGIRY QYNVLPQGWK GSPAIFQSSM TKILEPFRAK
ctsc1              ---------- ---------- ---------- ---------- --------tn
du422              ---------- ---------- ---------- --------rh a---------
du457              ---------- ---------- ---g------ --------rh a---------
GG5pol             ---------- ---------- ---------- ---------- ---------q
ctsc2              ---------- ---------- ---------- ---------- ---------q
du174              ---------- v------l-- ---------- ---------- ---------q
du151              ---------- ---------- ---------- --------a-- ----------
RB27pol            ---------- ---------- ---------- ---------- -r------tq
du204              ---------- ---------- ---------- ---------- --------t-
RB18pol            ---------- ---------- ---------- --------a-- --------tq
du156              ---------- v--------- ---------- ---------- ----------
du179              ---------- ---------- ---------- ---------- ---------q
GG3pol             ---------- --------x- ---------- --------a-- ----------
RB21pol            x--------- ---------- ---------- --------c-- ---------q
GG10pol            ---------- ---------- ---------- ---------- ----------
RB28pol            ---------- v------v-- ---------- ---------- -r--------
GG6pol             ---------- t--------- ---------- --------a-- -r------t-
RB13pol            ---------- t--------- ---------- --------c-- -r------t-
RB15pol            ---------- t------t-- ---------- ---------- -r------tq
GG4pol             ---------- ---a------ ---------- ---------- --------tq
RB22pol            ---------- f--------f ---------- ---------- ---------q
du172              ---------- ---------- ---gs----- ---s------ ----------
du115              ---------- ---aa--t-- ---------- ---------- i-------kn
RB14pol            ---------- ---------- ---------- --------a-- -r------tq
du368              ---------- ----p----- ---------- --------rh vr--------
du258              ---------- v------n-- -*-------- --------rh a---------q
Cpolcon            ---------- ---------- ---------- ---------- --------dr

                    151                                                200
sapep.msf(SApolcon) NPEIVIYQYM DDLYVGSDLE IGQHRAKIEE LREHLLKWGF TTPDKKHQKE
ctsc1              ---l------ ---------- ---------- ------r--- ----------
du422              ---------- ---------- ---------- --k---r--- ----------
du457              ---------- ---------- ---------- ---------- ----------
GG5pol             --g------- ---------- ---------- ---------- ----------
ctsc2              --g------- ---------- ---------- ------r--- ----------
du174              --g------- ---------- ---------- --d------- ----------
du151              ---------- ---------- ---------- --g------- ----------
RB27pol            --d------- ---------- ---------- ------r--- ----------
du204              --d------- ---------- ---------- ------r--- ----------
RB18pol            ---------- ---------- -------v-- ------r--- ----------
du156              ---------- ---------- ---------- ------r--- ----------
```

FIGURE 6 – continue

```
du179          ---------- ---------- ---------- ---------- ----------
GG3pol         ---------- ---------- ---------- ------r--- ----------
RB21pol        ---------- ---------- ---------- ---------- ----------
GG10pol        ---------- ---------- -r-------- ---------- ----------
RB28pol        ---------- ---------- ---------- --a------- ----------
GG6pol         --d------- ---------- ---------- ---------l ----------
RB13pol        ---------- ---------- ---------- ---------- ----------
RB15pol        ---------- ---------- -k-------- --a------- ----------
GG4pol         --d------- ---------- ---------- ------r--- ----------
RB22pol        ---l------ ---------- ------r--k ------r--- ----------
du172          --d--f---- ---------- -m-------d ---------- ----------
du115          --d------- ---------- ---------- ---------l ----------
RB14pol        ---------- ---------- -------v-- ------r--- ----------
du368          ---------- ---------- -------v-- --k------l ----------
du258          ---------- ---------- ---------k --d------l ----------
Cpolcon        ---l------ ---------- ---------- ---------- ----------

                       201                                        250
sapep.msf(SApolcon)    PPFLWMGYEL HPDKWTVQPI QLPEKDSWTV NDIQKLVGKL NWASQIYPGI
ctsc1          ---------- ---------- -----ed--- ---------- ----------
du422          ---------- ---------- ---------- ---------- ----------
du457          ---------- ---------- ---------- ---------- ----------
GG5pol         ---------- ---------- ---d------ ---------- -------a--
ctsc2          ---------- ---------- e--------- ---------- -------s--
du174          ---------- ---------- ---------- ---------- ----------
du151          ---------- ---------- ---------- ---------- ----------
RB27pol        ---------- ---------- ---------- ---------- ----------
du204          ---------- ---------- ---------- ---------- ----------
RB18pol        ---------- ---------- k--------- ---------- ----------
du156          ---------- ---------- ------d--- ---------- ----------
du179          ---------- ---------- n--d------ ---------- ----------
GG3pol         ---------- ---------- ---------- ---------- ----------
RB21pol        ---------- ---------- ---------- ---------- ----------
GG10pol        ---------- ---------- ---------- ---------- ----------
RB28pol        ---------- ---------- ---------- ---------- ----------
GG6pol         ---------- ---------- ---n------ ---------- -------s--
RB13pol        ---------- ---------- ---------- ---------- ----------
RB15pol        ---------- ---------- ---------- ---------- ----------
GG4pol         ---q------ ---------- c--------- ---------- ----------
RB22pol        ---------- ---------- -----e---- ---------- ----------
du172          ---------- ---------- ---------- ---------- ----------
du115          ---------- ---------- ---d------ ---------- ----------
RB14pol        ---------- ---------- ---------- ---------- ----------
du368          ---------- ---------- ---------- ---------- ----------
du258          ---------- ---------- --------x- ---------- ----------
Cpolcon        ---------- ---------- ---------- ---------- ----------
```

FIGURE 6 - continue

```
                 251                                 278
SApolcon)        KVRQLCKLLR GAKALTDIVP LTEEAELE
ctsc1            ---------- -t-------- --------
du422            --k------- ---------- --------
du457            ---------- ---------- --------
GG5pol           ---------- -------vi- --------
ctsc2            --------i- ---------- --------
du174            --k------- --------i- --------
du151            ---------- ---------- --------
RB27pol          ---------- --------i- --------
du204            ---------- ---------- --------
RB18pol          --k------- --------i- --------
du155            ---------- ---------- --------
du179            q--------- ---------- --------
GG3pol           ------r--- ---------- --------
RB21pol          ---------- -t-------- --------
GG10pol          --k-m----- -------vi- --------
RB28pol          ---------- ---------- --------
GG6pol           ---------- -t-------- --------
RB13pol          ---------- -t-----v-- --------
RB15pol          ---h------ -t-------- --------
GG4pol           ---h------ ---------- --------
RB22pol          --krm----- -------v-- --------
du172            ----m----- ---------- --------
du115            --kh--r--- ---------- --------
RB14pol          --k------- --------i- --------
du368            ---------- ---------a ------r-
du258            ---------- ---------- ----p---
Cpolcon          ---------- ---------- --------
```

FIGURE 7

Plurality: 2.00 Threshold: 4 AveWeight 1.00 AveMatch 2.91 AvMisMatch -2.00

PRETTY of: msf{*} May 1, 2000 11:45 ..

```
                 1                                                  50
SAenvcon         YCAPAGYAIL KCNNKTFNGT GPCNNVSTVQ CTHGIKPVVS TQLLLNGSLA
GG5env           ------f--- --kd------ ---------- ---------- ----------
du174env         ------f--- ---d------ ---------- ---------- ----------
RB13env          ---------- ---------- ---------- ---------- --f-------
du368env         ---------- ---------- ---h------ ---------- ----------
du422env         ---------- ---------- ---------- w--------- ----------
RB14env          ---------- ---------- ---h------ ---------- ----------
RB18env          ---------- ---------- ---y------ ---------- ----------
RB21env          ---------- ---------- ---------- ---------- ----------
GG6env           ---------- ---------- ---------- ---------- ----------
du123env         ---------- ---------- ---h------ ---------- ----------
du172env         ---------- ---------- ---------- ---------- ----------
du457env         ---------- ---------- ---h------ ---------- ----------
du151env         ---------- ---------- ---------- ---------- ----------
du467env         ---------- ---------- ---------- ---------- ----------
du179env         ---------- ---------- ---q------ ---------- --------i-
du204env         ---------- ---------- ---------- ---------- ----------
RB22env          ---------- -------d-- ---------- ---------- ----------
du258env         ---------- ----n----k ---t------ ---------- ----------
du281env         ---------- ----e----- ---------- ---------- ----------
RB12env          ---------- ---d-k---- ---y------ ---------- ----------
GG10env          ---------- ----e----- ---------l ---------- ----------
du115env         ---------- ---------s ---------- ---------- ----------
du156env         ---------- --td-k---- -s-------- ---------- ----------
GG4env           ---------- --k-e----- ---------- ---------- ----------
RB28env          ---------- ---------- ---------- ---------- ----------
GG3env           ------f--- ----q----- ---t------ ---------- ----------
RB27env          ---------- q--------- ---------- -----x---- ----------
Cenvcon          ---------- ---------- ---h------ ---------- ----------

                 51                                        100
Saenvcon         EEEIIIRSEN LTNNAKTIIV HLNESVEIVC TRPNNNTRKS IRIGPGQTFY
GG5env           kg----s-q- --d------- -----i--t- i-------q- ----------
du174env         -gg-----k- --d-s----- ----ti---- ---g----q- -------a-f
RB13env          -------k- --d--r---- ---------- --------q- -------a-f
du368env         -gkv----k- ----v----- ---k--n--- i--------g ----------
du422env         -----v---- ---s------ ---k----k- ---------- v---------
RB14en           -rd------- --d------- ---------- --------p- -------a--
RB18env          -rd------- --d------- ---------- --------q- -------a--
PB21env          ---------- ----v----- ---------- ---g------ v---------
GG6env           ---------- ----v----- -f-------- ---g------ ----------
du123env         ---------- ---------- -----i---- i--------- ----------
du172env         ---vv--f-- ------i--- ---------n- ---s------ v--------f
du457env         --d------- ---------- q--k------ ---------- ----------
du151env         ---------- ----i----- ---k------ ---------- ----------
du467env         -gk------- ---------- ---t--a--- ---------- ----------
du179env         -g-------- ----v----- -----ig--- ---g------ -------a--
du204env         --g------- ---------- q-----p--- i-------q- -------a-f
PB22env          ---------- ----v-i--- ---q--p-e- ---g------ v---------
du258env         -k--v----- ---------- q-enpi---- ---g------ v---------
du281env         --g---k--- m-d-i----- ---kl-k-e- ---s------ -------a-f
PB12env          ------k--d ----v----- -----ip--- i--g------ ----------
GG10env          k..t------ ------i--- ------q--- --------e- v---------
```

FIGURE 7 – continue

```
du115env      --k------- --d-t----i --t-----l- i--g------ -------i--
du156env      ------k--- --d-i----- q--q-ig-n- ---------- v---------
GG4env        -k-------- m-d-g----- ------r-e- i--------- v---------
RB28env       -g-------- ----i----- ------t-n- --------r- ----------
GG3env        -g----x--- --d-t----- ----p-a-n- ---g------ v---------
RB27env       -k-------- i---v----- ------q-t- ---h------ m---------
Cenvcon       ---------- ----v----- ---------- ---------- ----------

                101                                             150
SAenvcon  ATGDIIGDIR QAHCNISEGK WNKTLQKVKK KLKEELYKYK VVEIKPLGIA
GG5env    --k------- ----t----- --------ig ---------- ----------
du174env  --ke------ ---------q --------ae ----k----- ----------
RB13env   --kg------ --y----k-- --e-------i --g------- ----------
du368env  --na------ -------qa- --ta-kn--r --g-k----- --------v-
du422env  ---a------ e------ret --s--kq--g --g------- --------v-
RB14env   --h-------/e--------n -t----r-g- t-e---f--- ----------
RB18env   --h----n--/e--------n -t----r-g- t-e------- ----------
RB21env   ---------- -------sd- --q---q-g- --a------- ---v----v-
GG6env    ---e------ -------gan -t--m-r-s- ---------- --------l-
du123env  --n------- -------kt- --t--e---e ---d------ --------v-
du172env  ---------- -------re- --t---r--e ---------- ----------
du457env  --na------ --y----gad -----es--- --g------- --------v-
du151env  --da---n-- e------ksn -ts--eq--- ---------- ----------
du467env  --n----n-- --------eq -st-vaq--e --ra------ --------v-
du179env  .-nh------ --y----kqe -----ee-r- --q------- ----------
du204env  --k------- --y----..- --t--e---e r--------- --------v-
RB22env   ---------- --y--vt-er --i----ia- --lg------ ----------
du258env  ---------- ----t---e- --t-----ge --ik------ ----------
du281env  --na------ -------rdh -----e-i-g -f-------- --------v-
RB12env   -nn------- -------kcn -kl--v---* --hy------ --------v-
GG10env   +-------l- --p-s--in- ------e-s- --qk------ -i--------
du115env  ---g------ --y---n-ys -----kr-se -fr------- ---vr-----
du156env  ---------- -------rnq --e--eq--- --g------- --------v-
GG4env    ---qv----- --------rd --t---r-s- ---------- ----q---v-
RB28env   .-n------- -------rte --n--er-r- --e----l-t -------e--
GG3env    --dv-g-v-- a-r-dv-rxn --*-xeg--* --l------- ---v------
RB27env   ---v---i-q ppc-i-n-rx -wt-flh-gg e-l------- --------vv
Cenvcon   ---------- -------kd- --------s- --a------- --------v-

                151                                             200
SAenvcon  PTEAKRRVVE REKRAVGIGA VFLGFLGAAG STMGAASITL TVQARQLLSG
GG5env    --g-n----- ---------- m--------- ---------- ----------
du174env  --gt-w---- ------tl-- -------m-- ---------- ----------
RB13env   --t------- ------.--- m-------v-- -------l-- ------v---
du368env  --k------k ---------- -l-------- ---------- ----------
du422env  --ks--k--g -------l-- -l-------- ---------- ----------
RB14env   ---------- ---------- ---------- ---------- ----------
RB18env   ---------- ---------- ---------- ---------- ----------
RB21env   --a------- ---------- ---------- ---------- ----------
GG6env    ---------- ---------- ---------- ---------- ----------
du123env  --k------- ---------- -lf------- ---------- ----------
du172env  -dk------- ---------- ---------- -------m-- ----------
du457env  --k------- ---------- ---------- ---------- ----------
du151env  ---t------ ---------- ---------e- -------l-- ----------
du467env  ---s------ ----------- -l-------- -------v-- ----------
du179env  ---------- ---------- -l-------- ---------- ----------
du204env  --k------- ---------- ---------- ---------- p---------
RB22env   ---------- ---------- ------a--- ---------- ----------
du258env  --t------- ---------- ---------- ---------- ----------
du281env  ---------- ---------- -l-------- ---------- ----------
RB12env   ---------- .------m-- ---------- ---------- ----------
```

FIGURE 7 – continue

```
GG10env   --t------- ---------- m--------- ---------- ----------
du115env  --r------- -------a-- -if---a--- -------l-- ----------
du156env  --g---m--k -------l-- -lf------- ---------- -a--------
GG4env    --r------- ---------- al-------- -------ma- ----------
RB28env   -ik------- --i------- -l-------- ---------- ----------
GG3env    --ks------ -------m-- -if------- -------va- ----------
RB27env   ---------- ---------- ---------- ---------- ----------
Cenvcon   --k------- ---------- ---------- ---------- ----------

                201                          229
SAenvcon        IVQQQSNLLR AIEAQQHMLQ LTVWGIKQL
GG5env          ---------- ---------- ---------
du174env        ---------- ---------- ---------
RB13env         ---------k ---------- ---------
du368env        ---------- -------l-- ---------
du422env        ---------- -------l-- ---------
RB14env         ---------- ---------- ---------
RB18env         ---------- ---------- ---------
RB21env         ---------- ---v------ ---------
GG6env          ---------- ---------- ---------
du123env        ---------- ---------- ---------
du172env        ---------- ---------- ---------
du457env        ---------- ---------- ---------
du151env        ---------- ---------- --g------
du467env        ---------- ---------- ---------
du179env        ---------- ---------- ---------
du204env        ---------- -------l-- ---------
RB22env         -----n---- ---------- ---------
du258env        ---------- ---------- ---------
du281env        ---------- ---------- ---------
RB12env         ---------- ---------- ---------
GG10env         -----n---- ---------- ---------
du115env        ---------- ---------- ---------
du156env        ---------- ---------- ---------
GG4env          ---------- ----p----- -x-s-----
RB28env         v-------e-- --g-----m- -----v---
GG3env          -----n---- ---------- ---------
RB27env         ---------k -------l-- ---------
Cenvcon         ---------- ---------- ---------
```

FIGURE 8

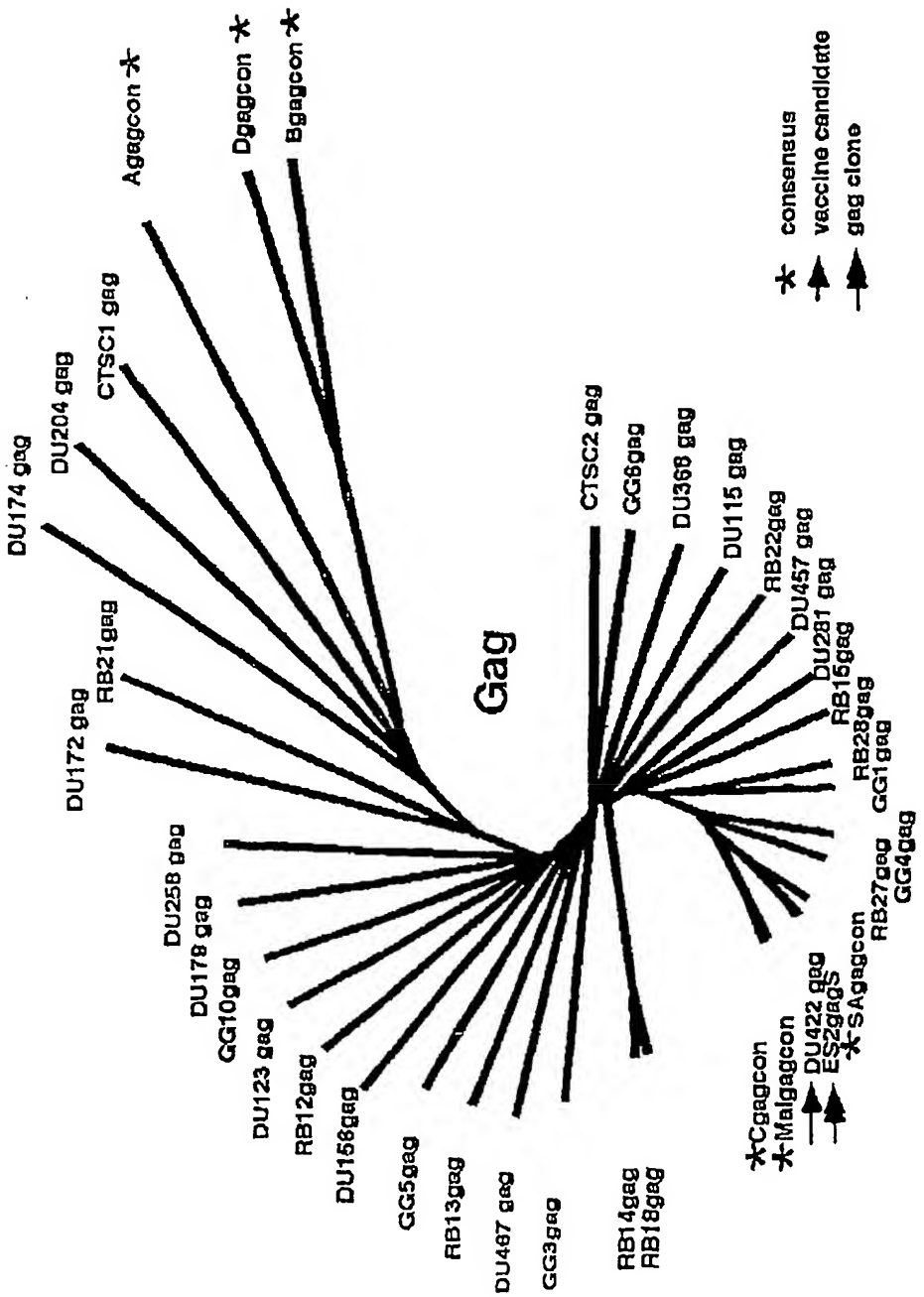

Gag
consensus
vaccine candidate
gag clone
Agagcon
Dgagcon
Bgagcon
CTSC1 gag
DU204 gag
DU174 gag
RB21gag
DU172 gag
DU258 gag
DU178 gag
GG10gag
DU123 gag
RB12gag
DU156gag
GG5gag
RB13gag
DU487 gag
GG3gag
RB14gag
RB18gag
CTSC2 gag
GG6gag
DU368 gag
DU115 gag
RB22gag
DU457 gag
DU281 gag
RB15gag
RB28gag
GG1gag
RB27gag
GG4gag
SAgagcon
DU422 gag
ES2gagS
Cgagcon
Malgagcon

FIGURE 9

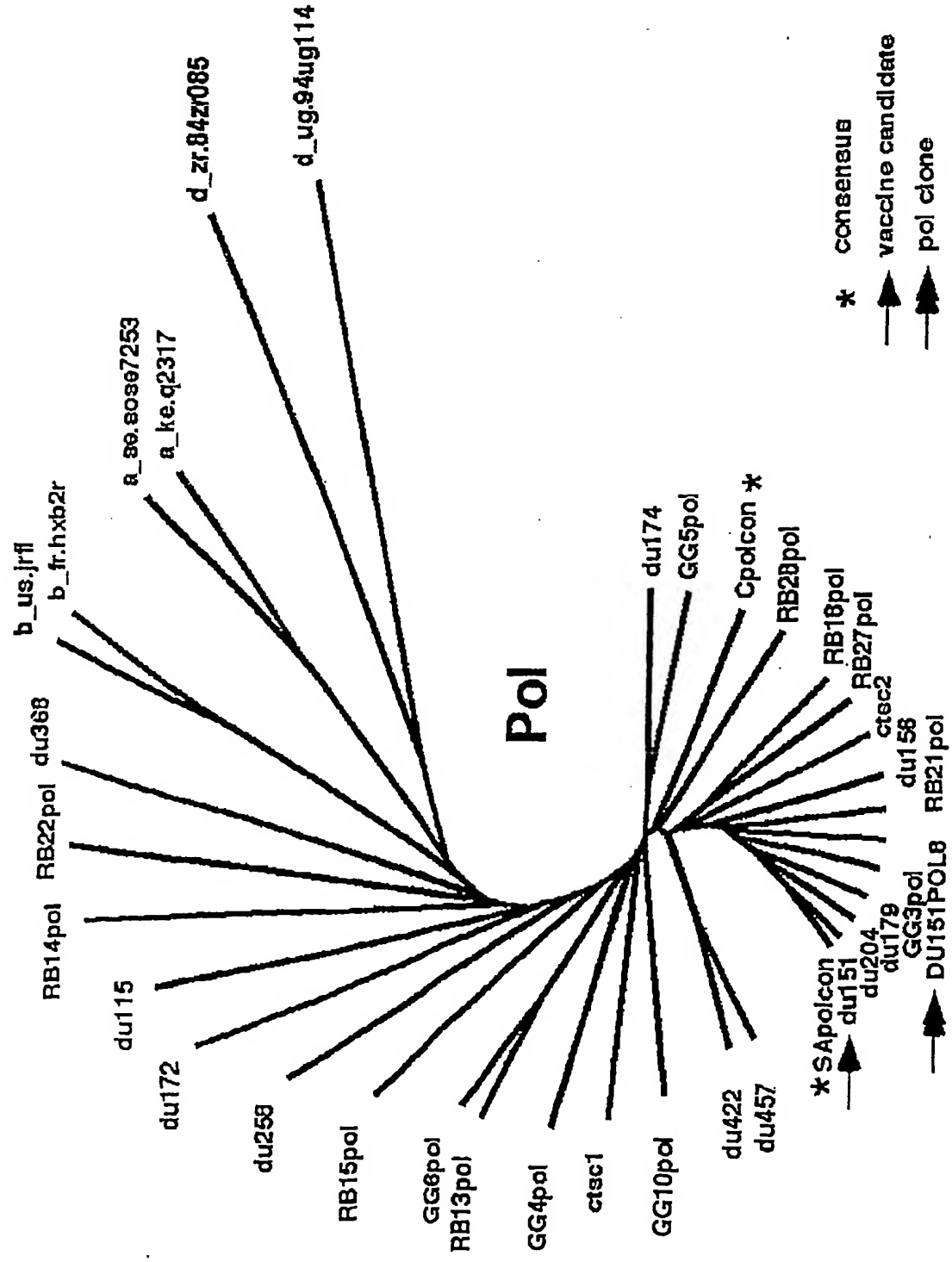

Pol
b_us.jrfl
b_fr.hxb2r
a_se.se0e7253
a_ke.q2317
d_zr.84zr085
d_ug.94ug114
RB14pol
RB22pol
du368
du115
du172
du258
RB15pol
GG6pol
RB13pol
GG4pol
ctsc1
GG10pol
du422
du457
*SApolcon
du151
du204
du179
GG3pol
DU151POL8
RB21pol
du156
ctsc2
RB27pol
RB18pol
RB28pol
Cpolcon *
GG5pol
du174
* consensus
vaccine candidate
pol clone

FIGURE 10

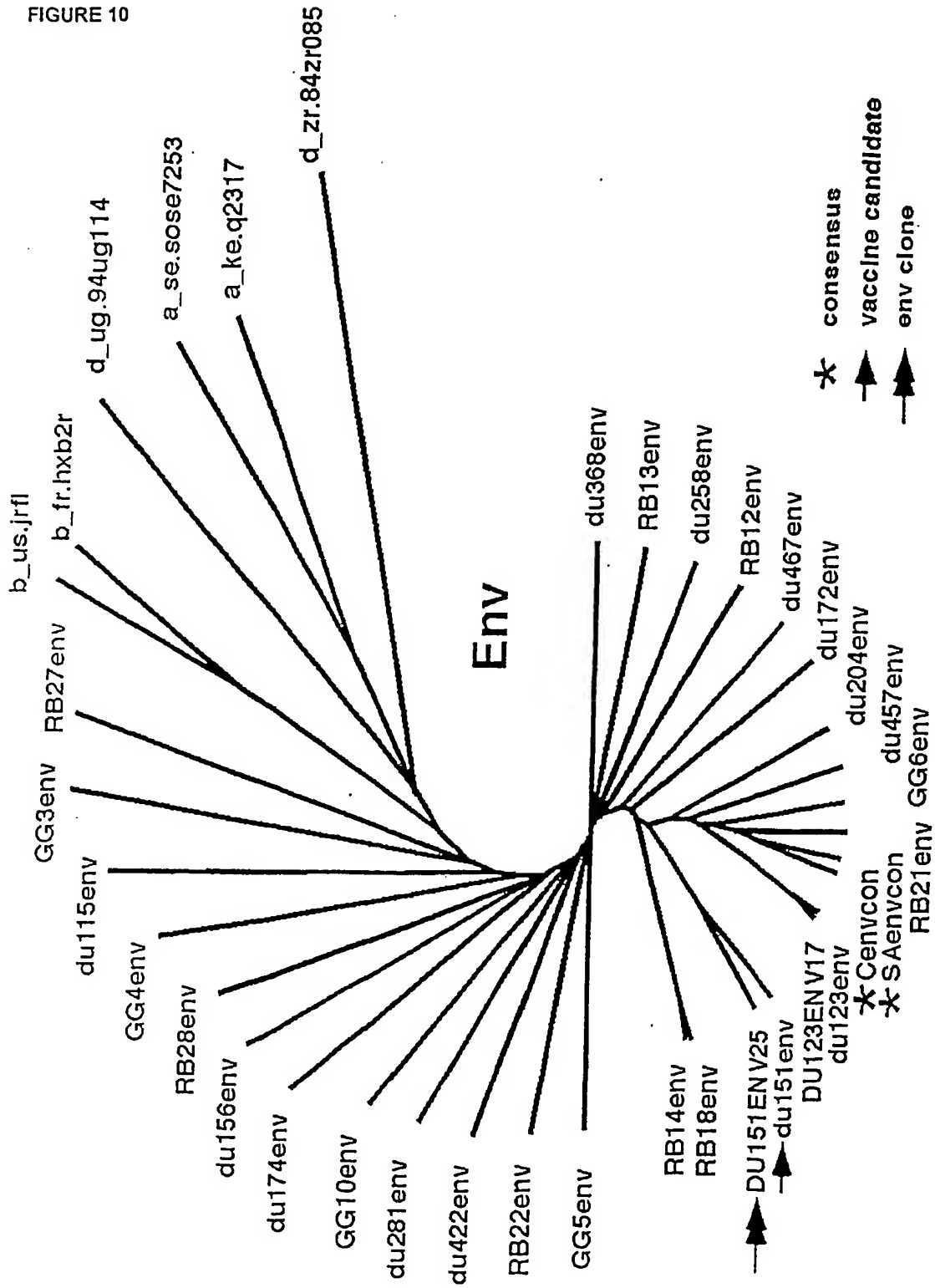

Env
b_us.jrfl
b_fr.hxb2r
d_ug.94ug114
a_se.sose7253
a_ke.q2317
d_zr.84zr085
GG3env
RB27env
du115env
GG4env
RB28env
du156env
du174env
GG10env
du281env
du422env
RB22env
GG5env
du368env
RB13env
du258env
RB12env
du467env
du172env
du204env
du457env
GG6env
RB21env
SAenvcon
Cenvcon
du123env
DU123ENV17
du151env
DU151ENV25
RB18env
RB14env
consensus
vaccine candidate
env clone

FIGURE 11

tha/IGAG lusta analysis
5/10/00

	DU115	DU123	DU156	DU172	DU174	DU179	DU204	DU258	DU281	DU368	DU422	gagclone	DU457	DU467	Sagagcon
DU115		88.9	92.2	89.1	88.6	90.9	88.9	92.5	92.8	92.2	95.1	94.8	92.8	91.2	95.4
DU123	88.9		90.2	87.1	85.4	88.9	88.6	89.9	90.6	91.2	91.2	91.2	90.2	89.6	92.5
DU156	92.2	90.2		89.1	89.6	92.5	89.3	91.9	92.2	91.2	94.8	94.8	96.9	90.6	96.1
DU172	89.0	87.1	89.1		88.1	89.1	88.4	90.7	90.7	89.7	90.4	90.4	90.4	88.4	92.0
DU174	88.6	85.4	89.6	88.1		90.0	88.7	90.3	88.7	88.3	91.3	90.6	90.3	89.0	91.6
DU179	90.9	88.9	92.5	89.1	90.0		89.3	92.2	90.9	92.1	92.2	92.2	90.6	91.2	93.5
DU204	88.9	88.6	89.3	88.4	88.7	89.3		90.6	89.9	88.8	90.6	90.6	88.9	87.9	91.2
DU258	92.5	89.6	91.9	90.7	90.3	92.2	90.6		90.9	89.9	92.9	92.9	91.2	90.9	94.2
DU281	92.8	90.6	92.2	90.7	88.7	90.9	89.9	90.9		94.8	98.4	98.4	94.5	93.8	95.8
DU368	92.2	91.2	91.2	89.7	88.3	91.2	88.9	89.8	94.8		98.1	95.8	94.8	94.1	95.1
DU422	95.1	91.2	94.8	90.4	91.3	92.2	90.6	92.9	88.4	98.1			98.1	95.1	98.0
DU457	92.8	90.2	91.9	90.4	90.3	90.6	88.8	91.2	94.5	94.8	98.1	95.8		95.1	95.8
DU467	91.2	89.6	90.6	88.4	89.0	91.2	87.9	90.9	93.8	94.1	95.1	94.5	95.1		93.8
AVE	91.3	89.3	91.3	89.3	89.0	90.8	89.2	91.2	92.2	91.9	93.5	93.3	92.7	91.4	94.2

FIGURE 12

hpl/POL fasta analysis
5/10/00

	DU115	DU151	poldone	DU156	DU172	DU174	DU179	DU204	DU258	DU281	DU368	DU422	DU457	Sapolcon
DU115		94.2	93.1	93.5	92.1	94.6	94.6	94.9	92.4	95.3	91.7	93.9	93.9	93.5
DU151	94.2			97.5	94.6	96.0	97.5	97.5	94.2	98.4	94.9	97.5	97.1	98.9
DU156	93.5	97.5	98.4		94.6	95.7	97.1	97.1	93.9	97.7	93.5	98.8	98.0	97.8
DU172	92.1	94.6	93.5	94.6		93.9	94.2	95.7	91.4	94.5	90.6	93.1	95.3	95.7
DU174	94.6	98.0	94.6	95.7	93.9		95.7	95.7	94.6	95.3	92.1	95.3	94.9	96.8
DU179	94.6	97.5	98.4	97.1	94.2	95.7		98.4	93.8	98.2	93.5	96.0	98.0	97.8
DU204	94.9	97.5	98.4	97.1	95.7	95.7	96.4		93.5	98.9	93.5	96.8	97.5	98.6
DU258	92.4	94.2	92.8	93.9	91.4	94.6	93.8	93.5		94.6	92.8	95.0	95.3	95.0
DU281	95.3	98.4	96.9	97.7	94.5	95.3	99.2	96.9	94.6		95.3	98.4	98.1	97.7
DU368	91.7	94.9	93.5	93.5	90.6	92.1	93.5	93.5	92.8	95.3		95.7	94.9	94.9
DU422	93.9	97.5	98.0	98.8	93.1	95.3	96.0	96.8	95.0	98.4	95.7		97.8	97.5
DU457	93.9	97.1	98.0	98.0	95.3	94.9	96.0	97.6	95.3	98.1	94.9	97.8		98.2
AVE	93.7	96.3	95.1	95.8	93.6	94.9	95.8	96.0	93.8	96.5	93.5	96.0	95.9	96.9

FIGURE 13

lhp//Env fasta analysis
5/10/00

	DU115	DU123	DU151	envelope	DU156	DU172	DU174	DU179	DU204	DU258	DU281	DU368	DU422	DU457	DU467	Saenvcon
DU115		85.6	83.4	83.1	81.2	84.3	82.5	84.7	83.8	83.0	83.4	82.1	82.1	85.2	83.8	87.3
DU123	85.6		89.5	90.8	87.8	90.4	85.8	90.4	92.1	89.5	88.2	90.4	87.8	92.6	90.8	93.9
DU151	83.4	89.6			85.2	87.8	84.3	88.2	87.3	85.2	87.3	88.0	89.1	90.8	89.5	92.8
DU156	81.2	87.8	85.2	87.3		85.2	82.5	85.2	84.3	83.8	85.6	83.8	87.8	87.3	85.2	87.6
DU172	84.3	90.4	87.8	87.3	85.2		84.7	86.0	89.1	89.5	86.5	86.0	87.3	88.8	87.8	93.0
DU174	82.5	85.6	84.3	84.3	82.5	84.7		88.0	86.5	84.7	85.2	83.0	81.7	84.7	84.3	89.1
DU179	84.7	90.4	88.2	91.3	85.2	88.0	86.0		87.8	86.5	87.3	86.8	84.7	90.4	87.8	91.3
DU204	83.8	82.1	87.3	87.8	84.3	89.1	86.5	87.8		86.5	87.8	87.8	85.6	91.3	87.8	91.7
DU258	83.0	88.5	85.2	84.7	83.8	89.5	84.7	86.5	86.5		83.4	83.8	84.3	87.8	86.9	91.3
DU281	83.4	88.2	87.3	87.3	85.6	86.5	85.2	87.3	87.8	83.4		85.6	86.5	89.1	86.5	89.5
DU368	82.1	90.4	86.0	86.0	83.8	86.0	83.0	86.9	87.8	83.8	85.6		87.8	90.8	88.2	88.6
DU422	82.1	87.8	89.1	91.7	87.8	87.3	81.7	84.7	85.6	84.3	86.5	87.8		89.1	87.3	89.1
DU457	85.2	92.6	90.8	90.8	87.3	88.8	84.7	90.4	91.3	87.8	89.1	90.8	89.1		89.5	93.4
DU467	83.8	90.8	89.5	90.0	85.2	87.8	84.3	87.8	87.8	86.9	86.5	88.2	87.3	89.5		91.3
AVE	83.5	89.3	87.2	87.9	85.0	87.2	84.3	87.1	87.5	85.8	86.3	86.3	86.2	89.0	87.3	90.7

FIGURE 14

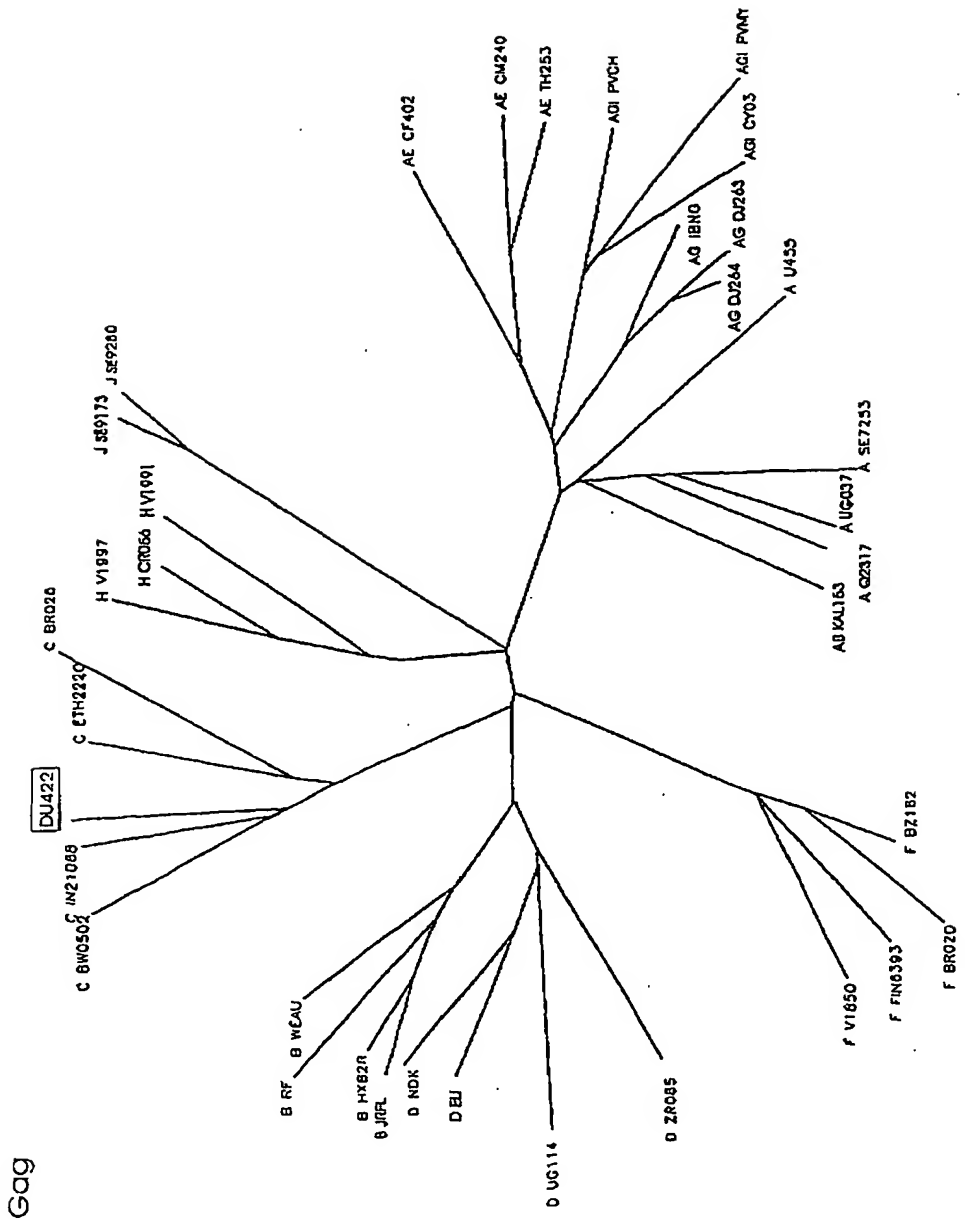

AE CF402
AE CM240
AE TH253
AGI PVCH
AGI PVMT
AGI CY03
AG IBNG
AG DJ243
AG DJ264
A U455
A SE7253
A UG037
A Q2317
AB KAL153
J SE9173
J SE9280
H CR056
H V1991
H V1997
C BR025
C ETH2220
DU422
C IN21088
C BW0502
B RF
B WEAU
B HXB2R
B JRFL
D NDK
D BJ
D UG114
D ZR085
F BZ162
F V1650
F FIN9363
F BR020
Gag

FIGURE 15

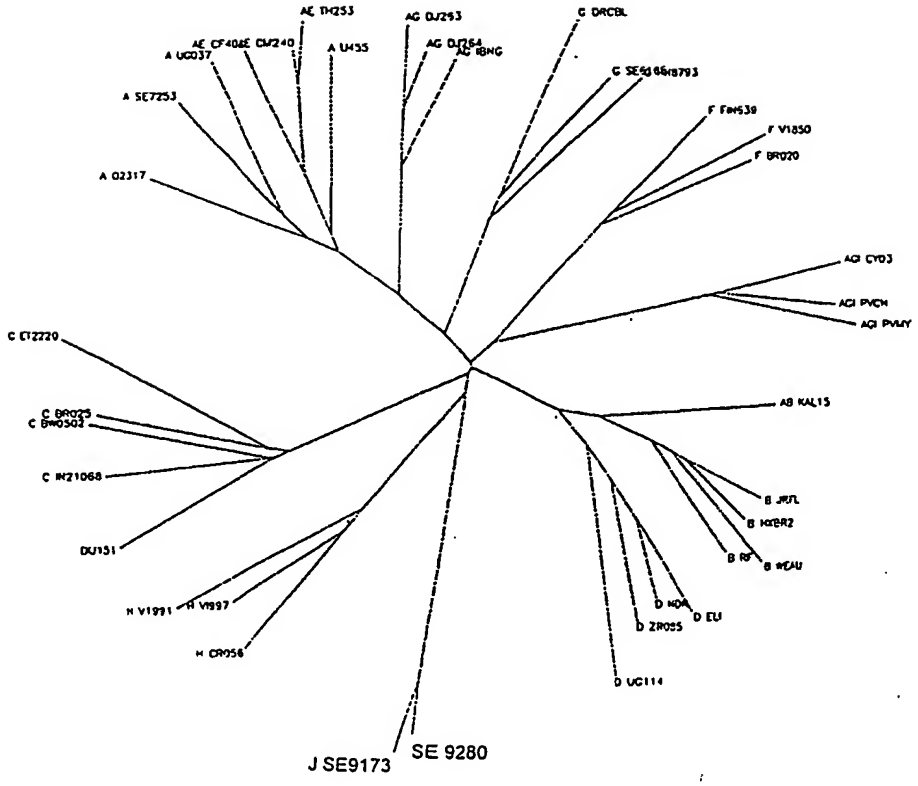
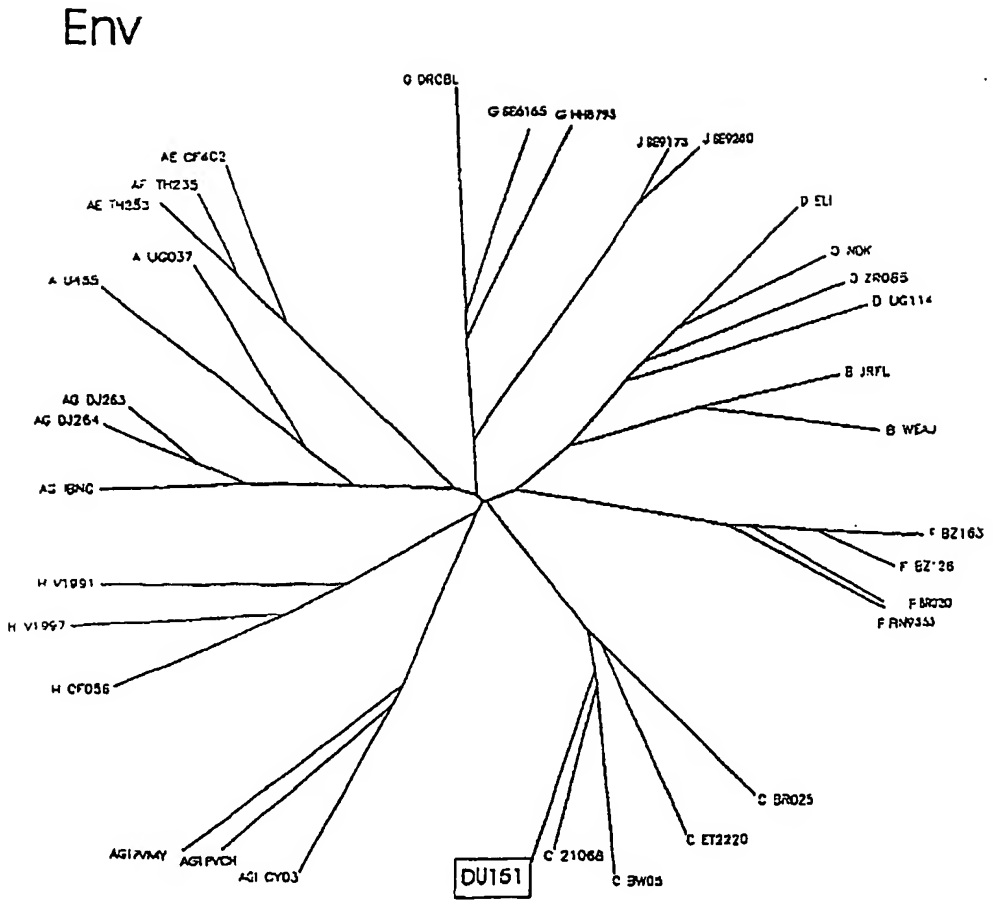

AE TH253
AG DJ263
G DRCBL
AE CF40AE DM240
A UG037
A UM55
AG DJ264
AG IBNG
G SE6366V8793
A SE7253
F FIM538
F V1850
F BR020
A 02317
AGI CY03
AGI PVCH
AGI PVMY
C ET2220
AB KAL15
C BR025
C BW0502
C IN21068
B JRFL
B HXB2R2
B RF
B WEAU
DU151
D NDK
D ELI
H V1991
H V997
D ZR055
H CR056
D UG114
J SE9173
SE 9280

FIGURE 16


Env
G DRCBL
G SE6165
G HH8793
J SE9173
J SE9280
AE CF4C2
AF TH235
AE TH253
A UG037
A U455
D ELI
D NDK
D ZR085
D UG114
B JRFL
B WEAU
AG DJ263
AG DJ264
AG IBNG
F BZ163
F BZ126
F BR020
F FIN9363
H VI991
H VI997
H CF056
C BR025
C ET2220
C 21068
C SW05
AGI AMMY
AGI PVCH
AGI CY03
DU151

SEQUENCE LISTING

SEQUENCE I.D. No 7: Du422 synthesised *gag* gene

```
1       GGGGGATGTGCTGCAAGGCGATTAAGTTGGGTAACGCCAGGGTTTTCCCAATCACGACGT
61      TGTAAAACGACAGCCAATGAATTGAAGCTTATGGCTGCTCGCGCATCTATCCTCAGAGGC
121     GAAAAGTTGGATAAGTGGGAAAAAATCAGACTCAGGCCAGGAGGTAAAAAACACTACATG
181     CTGAAGCATATCGTGTGGGCATCTAGGGAGTTGGAGAGATTTGCACTGAACCCCGGACTG
241     CTGGAAACCTCAGAGGGCTGTAAGCAAATCATGAAACAGCTCCAACCAGCCTTGCAGACC
301     GGAACAGAAGAGCTGAAGTCCCTTTACAATACCGTGGCAACCCTCTATTGCGTCCACGAG
361     AAGATCGAGGTGAGAGACACAAAGGAGGCCCTGGACAAAATCGAGGAGGAGCAGAATAAG
421     TGCCAGCAGAAGACCCAGCAGGCAAAGGCTGCTGACGGAAAGGTCTCTCAGAACTATCCT
481     ATCGTTCAGAACCTTCAGGGGCAGATGGTGCACCAAGCAATCAGCCCTAGAACCCTGAAC
541     GCATGGGTGAAGGTGATCGAGGAGAAAGCCTTTTCTCCCGAGGTTATCCCCATGTTTACC
601     GCCCTGAGCGAAGGCGCCACTCCTCAAGACCTGAACACTATGCTGAACACAGTGGGAGGA
661     CACCAGGCCGCTATGCAGATGTTGAAGGATACCATCAACGAGGAGGCAGCCGAATGGGAC
721     CGCCTCCACCCCGTGCACGCCGGACCTATCGCCCCCGGACAAATGAGAGAACCTCGCGGA
781     AGTGATATTGCCGGTACTACCAGCACCCTTCAAGAGCAGATTGCTTGGATGACCAGCAAC
841     CCACCCATCCCAGTGGGCGATATTTACAAAAGGTGGATTATTCTGGGGCTGAACAAAATT
901     GTGAGAATGTACTCCCCCGTCTCCATCCTCGACATCCGCCAAGGACCCAAGGAGCCTTTT
961     AGGGATTACGTGGACAGATTCTTCAAAACCCTTAGAGCTGAGCAAGCCACTCAGGAGGTT
1021    AAGAACTGGATGACAGATACTCTGCTCGTGCAAAACGCTAACCCCGATTGCAAAACCATC
1081    TTGAGAGCTCTCGGTCCAGGTGCCACCCTTGAGGAAATGATGACAGCATGTCAAGGCGTG
1141    GGAGGACCTGGGCACAAGGCCAGAGTTCTCGCTGAGGCCATGAGCCAGACAAACTCAGGC
1201    AATATCATGATGCAGAGGAGTAACTTTAAGGGTCCCAGGAGAATCGTCAAGTGCTTCAAT
1261    TGTGGCAAGGAGGGTCACATTGCCAGGAACTGCCGCGCCCCCAGGAAGAAAGGCTGCTGG
1321    AAGTGTGGCAAAGAGGGCCACCAGATGAAGGATTGCACCGAGCGCCAAGCAAACTTCCTG
1381    GGAAAGATTTGGCCCAGTCATAAGGGCCGCCCTGGCAACTTCCTTCAAAACAGACCCGAG
1441    CCTACCGCCCCCCCCGCTGAGTCTTTCAGATTTGAGGAGACCACCCCCGCTCCAAAGCAG
1501    GAGCCAATTGAGAGAGAGCCTCTCACCAGTCTCAAAAGCCTCTTTGGTAGCGACCCCCTC
1561    AGCCAATAAGAATTCTAGCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAATTG
1621    TTATCAGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGA
1681    TGCCTAATGAGTGAGCTAACTCACATTAGTTGCGTTGCGCTCACTGCCCGCTTTCCAGTC
1741    GGGAAACCTGTCGTGCCAGCTCCATTAGTGAATCGTCCAACGCACGGGGAGAGGCGGTTT
1801    GCGTATTGGGCGCACTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGTTCGTTCGGCT
1861    GCGGCGAGCCGTATCAGCTCACTCAAAGGCGGTAATACGGTTATC
```

SEQUENCE I.D. No 8: Du422 synthesised Gag Protein

```
1       GGGGGATGTGCTGCAAGGCGATTAAGTTGGGTAACGCCAGGGTTTTCCCAATCACGACGT
1        G  G  C  A  A  R  R  L  S  W  V  T  P  G  F  S  Q  S  R  R

61      TGTAAAACGACAGCCAATGAATTGAAGCTTATGGCTGCTCGCGCATCTATCCTCAGAGGC
21       C  K  T  T  A  N  E  L  K  L  M  A  A  R  A  S  I  L  R  G

121     GAAAAGTTGGATAAGTGGGAAAAAATCAGACTCAGGCCAGGAGGTAAAAAACACTACATG
41       E  K  L  D  K  W  E  K  I  R  L  R  P  G  G  K  K  H  Y  M

181     CTGAAGCATATCGTGTGGGCATCTAGGGAGTTGGAGAGATTTGCACTGAACCCCGGACTG
61       L  K  H  I  V  W  A  S  R  E  L  E  R  F  A  L  N  P  G  L

241     CTGGAAACCTCAGAGGGCTGTAAGCAAATCATGAAACAGCTCCAACCAGCCTTGCAGACC
81       L  E  T  S  E  G  C  K  Q  I  M  K  Q  L  Q  P  A  L  Q  T

301     GGAACAGAAGAGCTGAAGTCCCTTTACAATACCGTGGCAACCCTCTATTGCGTCCACGAG
101      G  T  E  E  L  K  S  L  Y  N  T  V  A  T  L  Y  C  V  H  E

361     AAGATCGAGGTGAGAGACACAAAGGAGGCCCTGGACAAAATCGAGGAGGAGCAGAATAAG
121      K  I  E  V  R  D  T  K  E  A  L  D  K  I  E  E  E  Q  N  K
```

```
421   TGCCAGCAGAAGACCCAGCAGGCAAAGGCTGCTGACGGAAAGGTCTCTCAGAACTATCCT
141    C  Q  Q  K  T  Q  Q  A  K  A  A  D  G  K  V  S  Q  N  Y  P

481   ATCGTTCAGAACCTTCAGGGGCAGATGGTGCACCAAGCAATCAGCCCTAGAACCCTGAAC
161    I  V  Q  N  L  Q  G  Q  M  V  H  Q  A  I  S  P  R  T  L  N

541   GCATGGGTGAAGGTGATCGAGGAGAAAGCCTTTTCTCCCGAGGTTATCCCCATGTTTACC
181    A  W  V  K  V  I  E  E  K  A  F  S  P  E  V  I  P  M  F  T

601   GCCCTGAGCGAAGGCGCCACTCCTCAAGACCTGAACACTATGCTGAACACAGTGGGAGGA
201    A  L  S  E  G  A  T  P  Q  D  L  N  T  M  L  N  T  V  G  G

661   CACCAGGCCGCTATGCAGATGTTGAAGGATACCATCAACGAGGAGGCAGCCGAATGGGAC
221    H  Q  A  A  M  Q  M  L  K  D  T  I  N  E  E  A  A  E  W  D

721   CGCCTCCACCCCGTGCACGCCGGACCTATCGCCCCCGGACAAATGAGAGAACCTCGCGGA
241    R  L  H  P  V  H  A  G  P  I  A  P  G  Q  M  R  E  P  R  G

781   AGTGATATTGCCGGTACTACCAGCACCCTTCAAGAGCAGATTGCTTGGATGACCAGCAAC
261    S  D  I  A  G  T  T  S  T  L  Q  E  Q  I  A  W  M  T  S  N

841   CCACCCATCCCAGTGGGCGATATTTACAAAAGGTGGATTATTCTGGGGCTGAACAAAATT
281    P  P  I  P  V  G  D  I  Y  K  R  W  I  I  L  G  L  N  K  I

901   GTGAGAATGTACTCCCCCGTCTCCATCCTCGACATCCGCCAAGGACCCAAGGAGCCTTTT
301    V  R  M  Y  S  P  V  S  I  L  D  I  R  Q  G  P  K  E  P  F

961   AGGGATTACGTGGACAGATTCTTCAAAACCCTTAGAGCTGAGCAAGCCACTCAGGAGGTT
321    R  D  Y  V  D  R  F  F  K  T  L  R  A  E  Q  A  T  Q  E  V

1021  AAGAACTGGATGACAGATACTCTGCTCGTGCAAAACGCTAACCCCGATTGCAAAACCATC
341    K  N  W  M  T  D  T  L  L  V  Q  N  A  N  P  D  C  K  T  I

1081  TTGAGAGCTCTCGGTCCAGGTGCCACCCTTGAGGAAATGATGACAGCATGTCAAGGCGTG
361    L  R  A  L  G  P  G  A  T  L  E  E  M  M  T  A  C  Q  G  V

1141  GGAGGACCTGGGCACAAGGCCAGAGTTCTCGCTGAGGCCATGAGCCAGACAAACTCAGGC
381    G  G  P  G  H  K  A  R  V  L  A  E  A  M  S  Q  T  N  S  G

1201  AATATCATGATGCAGAGGAGTAACTTTAAGGGTCCCAGGAGAATCGTCAAGTGCTTCAAT
401    N  I  M  M  Q  R  S  N  F  K  G  P  R  R  I  V  K  C  F  N

1261  TGTGGCAAGGAGGGTCACATTGCCAGGAACTGCCGCGCCCCCAGGAAGAAAGGCTGCTGG
421    C  G  K  E  G  H  I  A  R  N  C  R  A  P  R  K  K  G  C  W

1321  AAGTGTGGCAAAGAGGGCCACCAGATGAAGGATTGCACCGAGCGCCAAGCAAACTTCCTG
441    K  C  G  K  E  G  H  Q  M  K  D  C  T  E  R  Q  A  N  F  L

1381  GGAAAGATTTGGCCCAGTCATAAGGGCCGCCCTGGCAACTTCCTTCAAAACAGACCCGAG
461    G  K  I  W  P  S  H  K  G  R  P  G  N  F  L  Q  N  R  P  E

1441  CCTACCGCCCCCCCCCGCTGAGTCTTTCAGATTTGAGGAGACCACCCCCGCTCCAAAGCAG
481    P  T  A  P  P  A  E  S  F  R  F  E  E  T  T  P  A  P  K  Q

1501  GAGCCAATTGAGAGAGAGCCTCTCACCAGTCTCAAAAGCCTCTTTGGTAGCGACCCCCTC
501    E  P  I  E  R  E  P  L  T  S  L  K  S  L  F  G  S  D  P  L

1561  AGCCAATAAGAATTCTAGCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAATTG
521    S  Q  *  E  F  *  L  G  V  I  M  V  I  A  V  S  C  V  K  L

1621  TTATCAGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGA
541    L  S  A  H  N  S  T  Q  H  T  S  R  K  H  K  V  *  S  L  G
```

```
1681    TGCCTAATGAGTGAGCTAACTCACATTAGTTGCGTTGCGCTCACTGCCCGCTTTCCAGTC
561      C  L  M  S  E  L  T  H  I  S  C  V  A  L  T  A  R  F  P  V

1741    GGGAAACCTGTCGTGCCAGCTCCATTAGTGAATCGTCCAACGCACGGGGAGAGGCGGTTT
581      G  K  P  V  V  P  A  P  L  V  N  R  P  T  H  G  E  R  R  F

1801    GCGTATTGGGCGCACTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGTTCGTTCGGCT
601      A  Y  W  A  H  F  R  F  L  A  H  *  L  A  A  L  V  R  S  A

1861    GCGGCGAGCCGTATCAGCTCACTCAAAGGCGGTAATACGGTTATC
621      A  A  S  R  I  S  S  L  K  G  G  N  T  V  I
```

SEQUENCE I.D. No 9: Du151 synthesised *pol* gene

```
1       TCGCGCGTTT CGGTGATGAC GGTGAAAACC TCTGACACAT GCAGCTCCCG
51      GAGACGGTCA CAGCTTGTCT GTAAGCGGAT GCCGGGAGCA GACAAGCCCG
101     TCAGGGCGCG TCAGCGGGTG TTGGCGGGTG TCGGGGCTGG CTTAACTATG
151     CGGCATCAGA GCAGATTGTA CTGAGAGTGC ACCATATGCG GTGTGAAATA
201     CCGCACAGAT GCGTAAGGAG AAAATACCGC ATCAGGCGCC ATTCGCCATT
251     CAGGCTGCGC AACTGTTGGG AAGGGCGATC GGTGCGGGCC TCTTCGCTAT
301     TACGCCAGCT GGCGAAAGGG GGATGTGCTG CAAGGCGATT AAGTTGGGTA
351     ACGCCAGGGT TTTCCCAGTC ACGACGTTGT AAAACGACGG CCAGTGCCAA
401     GCTTGCATGC CTGCAGGTCG ACTCTAGAGG ATCCCCGGGT ACCGAGCTCC
                    BglI (join to Gag for Gag-pol)
                  ~~~~~~~~~~~~
451     TTCCCACAAG GGCCGGCCAG GCAATTTCCT TCAGAACAGA CCAGAGCCAA
501     CAGCCCCACC AGCAGAGAGC TTCAGGTTCG AAGAGACAAC CCCCGCTCCG
551     AAACAGGAGC CGAGAGAAAG GGAACCCTTA ACTTCCCTCA AATCACTCTT
601     TGGCAGCGAC CCCTTGTCTC AATAAAAATC GGCGGCCAGA CCCGGGAGGC
651     CCTGCTGGAC ACCGGCGCCG ACGACACCGT GCTGGAGGAC ATCAACCTGC
701     CCGGCAAGTG GAAGCCCAAG ATGATCGGCG GCATCGGCGG CTTCATCAAG
751     GTGCGGCAGT ACGACCAGAT CCTGATCGAG ATCTGCGGCA AGAAGGCCAT
801     CGGCACCGTG CTGGTGGGCC CCACCCCCGT GAACATCATC GGCCGGAACA
851     TGCTGACCCA GCTGGGCTGC ACCCTGAACT TCCCCATCAG CCCCATCGAG
901     ACCGTGCCCG TGAAGCTGAA GCCCGGCATG GACGGCCCCA AGGTGAAGCA
951     GTGGCCCCTG ACCGAGGTGA AGATCAAGGC CCTGACCGCC ATCTGCGAGG
1001    AGATGGAGAA GGAGGGCAAG ATCACCAAGA TCGGCCCCGA GAACCCCTAC
1051    AACACCCCCA TCTTCGCCAT CAAGAAGGAG GACAGCACCA AGTGGCGGAA
1101    GCTGGTGGAC TTCCGGGAGC TGAACAAGCG GACCCAGGAC TTCTGGGAGG
1151    TGCAGCTGGG CATCCCCCAC CCCGCCGGCC TGAAGAAGAA GAAGAGCGTG
1201    ACCGTGCTGG ACGTGGGCGA CGCCTACTTC AGCGTGCCCC TGGACGAGGG
1251    CTTCCGGAAG TACACCGCCT TCACCATCCC CAGCATCAAC AACGAGACCC
1301    CCGGCATCCG GTACCAGTAC AACGTGCTGC CCCAGGGCTG GAAGGGCAGC
1351    CCCGCCATCT TCCAGGCCAG CATGACCAAG ATCCTGGAGC CCTTCCGGGC
1401    CAAGAACCCC GAGATCGTGA TCTACCAGTA CATGGCCGCC CTGTACGTGG
1451    GCAGCGACCT GGAGATCGGC CAGCACCGGG CCAAGATCGA GGAGCTGCGG
1501    GAGCACCTGC TGAAGTGGGG CTTCACCACC CCCGACAAGA AGCACCAGAA
1551    GGAGCCCCCC TTCCTGTGGA TGGGCTACGA GCTGCACCCC GACAAGTGGA
1601    CCGTGCAGCC CATCCAGCTG CCCGAGAAGG ACAGCTGGAC CGTGAACGAC
1651    ATCCAGAAGC TGGTGGGCAA GCTGAACTGG ACCAGCCAGA TCTACCCCGG
1701    CATCAAGGTG CGGCAGCTGT GCAAGCTGCT GCGGGGCACC AAGGCCCTGA
1751    CCGACATCGT GCCCCTGACC GAGGAGGCCG AGCTGGAGCT GGCCGAGAAC
1801    CGGGAGATCC TGAAGGAGCC CGTGCACGGC GTGTACTACG ACCCCAGCAA
1851    GGACCTGATC GCCGAGATCC AGAAGCAGGG CGACGACCAG TGGACCTACC
```

```
1901    AGATCTACCA GGAGCCCTTC AAGAACCTGA AAACCGGCAA GTACGCCAAG
1951    CGGCGGACCA CCCACACCAA CGACGTGAAG CAGCTGACCG AGGCCGTGCA
2001    GAAGATCAGC CTGGAGAGCA TCGTGACCTG GGGCAAGACC CCCAAGTTCC
2051    GGCTGCCCAT CCAGAAGGAG ACCTGGGAGA TCTGGTGGAC CGACTACTGG
2101    CAGGCCACCT GGATCCCCGA GTGGGAGTTC GTGAACACCC CCCCCCTGGT
2151    GAAGCTGTGG TACCAGCTGG AGAAGGAGCC CATCGCCGGC GCCGAGACCT
2201    TCTACGTGGA CGGCGCCGCC AACCGGGAGA CCAAGATCGG CAAGGCCGGC
2251    TACGTGACCG ACCGGGGCCG GCAGAAGATC GTGACCCTGA GCGAGACCAC
2301    CAACCAGAAA ACCGAGCTGC AGGCCATCCA GCTGGCCCTG CAGGACAGCG
2351    AGAGCGAGGT GAACATCGTG ACCGACAGCC AGTACGCCCT GGGCATCATC
2401    CAGGCCCAGC CCGACCGGAG CGAGAGCGAG CTGGTGAACC AGATCATCGA
2451    GCAGCTGATC AAGAAGGAGC GGGCCTACCT GAGCTGGGTG CCCGCCCACA
2501    AGGGCATCGG CGGCGACGAG CAGGTGGACA AGCTGGTGAG CAGCGGCATC
2551    CGGAAGGTGC TGTGATCTAG AGAATTC
```

SEQUENCE I.D. No 10: Du151 synthesised Pol Protein

```
1       SRVSVMTVKT SDTCSSRRRS QLVCKRMPGA DKPVRARQRV LAGVGAGLTM RHQSRLY*EC
61      TICGVKYRTD A*GENTASGA IRHSGCATVG KGDRCGPLRY YASWRKGDVL QGD*VG*RQG
121     FPSHDVVKRR PVPSLHACRS TLEDPRVPSS FPQGPARQFP SEQTRANSPT SRELQVRRDN
181     PRSETGAERK GTLNFPQITL WQRPLVSIKI GGQTREALLD TGADDTVLED INLPGKWKPK
241     MIGGIGGFIK VRQYDQILIE ICGKKAIGTV LVGPTPVNII GRNMLTQLGC TLNFPISPIE
301     TVPVKLKPGM DGPKVKQWPL TEVKIKALTA ICEEMEKEGK ITKIGPENPY NTPIFAIKKE
361     DSTKWRKLVD FRELNKRTQD FWEVQLGIPH PAGLKKKKSV TVLDVGDAYF SVPLDEGFRK
421     YTAFTIPSIN NETPGIRYQY NVLPQGWKGS PAIFQASMTK ILEPFRAKNP EIVIYQYMAA
481     LYVGSDLEIG QHRAKIEELR EHLLKWGFTT PDKKHQKEPP FLWMGYELHP DKWTVQPIQL
541     PEKDSWTVND IQKLVGKLNW TSQIYPGIKV RQLCKLLRGT KALTDIVPLT EEAELELAEN
601     REILKEPVHG VYYDPSKDLI AEIQKQGDDQ WTYQIYQEPF KNLKTGKYAK RRTTHTNDVK
661     QLTEAVQKIS LESIVTWGKT PKFRLPIQKE TWEIWWTDYW QATWIPEWEF VNTPPLVKLW
721     YQLEKEPIAG AETFYVDGAA NRETKIGKAG YVTDRGRQKI VTLSETTNQK TELQAIQLAL
781     QDSESEVNIV TDSQYALGII QAQPDRSESE LVNQIIEQLI KKERAYLSWV PAHKGIGGDE
841     QVDKLVSSGI RKVL*
```

SEQUENCE I.D. No 11: Du151 synthesised *env* Gene

```
1       AAGCTTATGA GGGTTATGGG GATTCAGAGA AACTGGCCTC AGTGGTGGAT TTGGGGGACA
61      TTGGGATTTT GGATGATCAT CATCTGTCGC GTCGTGGGCA ACCTGAACCT GTGGGTCACT
121     GTCTACTATG GAGTGCCAGT TTGGAAGGAA GCCAAGACAA CTCTGTTTTG CGCCAGCGAC
181     GCCAAGGCTT ATGACAAGGA AGTCCACAAC GTGTGGGCCA CCCACGCATG TGTCCCAACC
241     GACCCCAACC CACGCGAAAT CGTGCTGGAA AACGTCACAG AAAATTTCAA CATGTGGAAA
301     AACGATATGG TGGATCAGAT GCATGAGGAT ATTATTAGCC TCTGGGACCA GTCTCTGAAG
361     CCATGTGTGA AGTTGACACC TCTCTGTGTG ACCCTTAACT GTACTAACGC CCCCGCCTAT
421     AACAACTCTA TGCACGGGGA GATGAAAAAC TGTTCCTTCA ACACCACCAC CGAAATCAGG
481     GACAGAAAAC AGAAAGCCTA TGCCCTGTTC TATAAGCCCG ATGTGGTGCC ACTTAACCGC
541     CGCGAAGAAA ATAATGGTAC TGGCGAATAT ATTCTGATTA ACTGTAACAG CTCTACAATT
601     ACTCAGGCTT GCCCTAAAGT CACCTTTGAC CCAATCCCAA TCCACTACTG CGCCCCTGCA
661     GGATACGCTA TCCTGAAATG CAATAATAAG ACCTTCAACG GAACTGGACC CTGCAATAAC
721     GTGTCTACAG TGCAATGTAC CCACGGCATT ATGCCCGTCG TCTCCACCCA ACTGCTGCTC
781     AATGGCAGCT TGGCAGAAGA GGAGATCATT ATTAGGAGCG AAAACCTCAC CAACAATATC
841     AAGACAATCA TCGTGCACCT GAACAAGTCT GTGGAAATTG TGTGTACCAG GCCCAATAAC
901     AACACCAGGA AGAGCATCCG CATCGGACCT GGACAAACTT TCTACGCCAC CGGCGAAATC
961     ATCGGGAACA TTAGAGAAGC CCACTGCAAC ATCTCTAAGA GCAATTGGAC ATCTACATTG
1021    GAGCAAGTGA AAAAAAAGCT GAAAGAGCAC TACAATAAGA CCATCGAGTT CAACCCTCCT
1081    TCCGGCGGCG ATCTGGAGGT CACAACACAC TCCTTTAACT GTAGGGGGGA GTTCTTTTAC
1141    TGCAACACAA CAAAGCTGTT TAGCAACAAC TCCGACAGCA ATAATGAGAC TATCACCCTG
1201    CCTTGCAAGA TCAAGCAAAT CATTAACATG TGGCAGAAAG TGGGAAGGGC AATGTATGCA
1261    CCTCCCATCG AGGGCAACAT CACATGCAAG TCTAATATCA CCGGCCTGTT GCTGACTAGA
1321    GACGGTGGCA AGAATACTAC TAACGAAATC TTCAGGCCAG GTGGAGGGAA CATGAAAGAT
```

```
1381   AATTGGCGCT CCGAACTGTA TAAGTACAAG GTGGTGGAGA TTGAGCCCCT CGGCGTCGCC
1441   CCCACAAAGT CTAAGCGCCG CGTGGTGGAA AGAGAGAAGA GGGCTGTCGG CCTCGGCGCA
1501   GTGCTGCTGG GGTTCTTGGG TGCCGCTGGG TCTACAATGG GCGCTGCCTC TATTACACTC
1561   ACCGTGCAAG CTAGGCAGCT GCTGTCCGGT ATTGTGCAAC AACAGAGCAA TCTCTTGAGA
1621   GCTATCGAGG CCCAGCAGCA TATGCTGCAA CTTACAGTGT GGGGTATTAA GCAGCTGCAA
1681   ACTCGCGTCC TGGCAATCGA ACGCTACCTG AAAGACCAGC AACTCCTGGG TCTGTGGGGC
1741   TGCTCCGGTA AGATCATCTG TACCACAGCC GTGCCCTGGA ACACCAGCTG GTCCAATAAG
1801   AGCCAAGAGG ATATTTGGGA TAATATGACC TGGATGCAAT GGGATAGAGA GATCAGCAAC
1861   TACACAGGAA CCATTTATAG GCTCCTGGAA GATTCTCAGA ACCAGCAGGA GAAGAACGAG
1921   AAGGACTTGC TCGCCCTGGA TAGCTGGAAA AACCTGTGGA ATTGGTTTAA CATCACCAAC
1981   TGGCTTTGGT ACATTAAGAT TTTCATCATG ATTGTGGGAG GCTTGATCGG CCTGAGGATT
2041   ATCTTCGGGG TGCTTGCCAT TGTGAAAAGG GTCAGACAAG GATACTCCCC ATTGTCCTTT
2101   CAGACCTTGA CTCCAAGCCC ACGCGGACCC GACAGGTTGG GCAGGATCGA GGAGGAAGGA
2161   GGCGAACAGG ATAAGGACCG CTCCATCAGA CTTGTTAGCG GGTTTCTGGC CCTGGCCTGG
2221   GATGATCTGA GGAGCCTGTG CCTCTTCTCC TATCACCACC TCCGCGATTT CATCCTCATT
2281   GCAGCTAGGG CTGCTGAGTT GCTGGGACGC TCCTCCCTGA GAGGTCTCCA GAGAGGCTGG
2341   GAGGCACTGA AGTACCTCGG GAACCTTGTG CAATACGGCG GGCTGGAGCT GAAAAGATCC
2401   GCCATCAAGC TGTTCGACAC CATCGCAATC GCCGTTGCAG AGGGCACCGA CAGGATCTTG
2461   GAGGTCATTC AGAGGATCTG TCGCGCCATC CGCCACATCC CCATCAGGAT CAGACAAGGA
2521   TTCGAGGCAG CACTGCAATG ATAGTTAATT AAACGCGTGG ATCC
```

SEQUENCE I.D. No 12: Du151 synthesised Env Protein

```
KLMRVMGIQRNWPQWWIWGTLGFWMIIICRVVGNLNLWVTVYYGVPVWKEAKTTLFCASD
AKAYDKEVHNVWATHACVPTDPNPREIVLENVTENFNMWKNDMVDQMHEDIISLWDQSLK
PCVKLTPLCVTLNCTNAPAYNNSMHGEMKNCSFNTTTEIRDRKQKAYALFYKPDVVPLNR
REENNGTGEYILINCNSSTITQACPKVTFDPIPIHYCAPAGYAILKCNNKTFNGTGPCNN
VSTVQCTHGIMPVVSTQLLLNGSLAEEEIIIRSENLTNNIKTIIVHLNKSVEIVCTRPNN
NTRKSIRIGPGQTFYATGEIIGNIREAHCNISKSNWTSTLEQVKKKLKEHYNKTIEFNPP
SGGDLEVTTHSFNCRGEFFYCNTTKLFSNNSDSNNETITLPCKIKQIINMWQKVGRAMYA
PPIEGNITCKSNITGLLLTRDGGKNTTNEIFRPGGGNMKDNWRSELYKYKVVEIEPLGVA
PTKSKRRVVEREKRAVGLGAVLLGFLGAAGSTMGAASITLTVQARQLLSGIVQQQSNLLR
AIEAQQHMLQLTVWGIKQLQTRVLAIERYLKDQQLLGLWGCSGKIICTTAVPWNSSWSNK
SQEDIWDNMTWMQWDREISNYTGTIYRLLEDSQNQQEKNEKDLLALDSWKNLWNWFNITN
WLWYIKIFIMIVGGLIGLRIIFGVLAIVKRVRQGYSPLSFQTLTPSPRGPDRLGRIEEEG
GEQDKDRSIRLVSGFLALAWDDLRSLCLFSYHHLRDFILIAARAAELLGRSSLRGLQRGW
EALKYLGNLVQYGGLELKRSAIKLFDTIAIAVAEGTDRILEVIQRICRAIRHIPIRIRQG
FEAALQOOLIKRVD*
```

SEQUENCE I.D. No 13: Du179 Env Gene (non-humanised)

```
AGGCTAATTTTTTAGGGAAAATTTGGCCTTCCCACAAGGGGAGGCCAGGGAATTTCCTTCAGAGCAGGCCAATGAGAGT
GAGGGGGATACAGAGGAATTGGCCACAATGGTGGATATGGGGCATCTTAGGCTTTTGGATGTTAATGATTTGTAGTGGG
GTGGGAAACTTGTGGGTCACAATCTATTATGGGGTACCTGTGTGGAGAGAAGCAAAAACTACTCTATTCTGTGCATCAG
ATGCTAAAGCATATGATAGAGAAGTGCATAATGTCTGGGCTACACATGCCTGTGTACCCACAGACCCCAACCCACAAGA
AATAGTTATGGGAAATGTAACAGAAAATTTTAACATGTGGAAAAATGACATGGTGGATCAGATGCATGAGGATATAATC
AATTTATGGGATCAAAGCCTAAAGCCATGTGTAAAGTTAACCCCACTCTGTGTCACTTTAAAATGTAGTACCTATAATG
GTAGTGATACCAACGATATGAGAAATTGCTCTTTCAATACAACTACAGAAATAAGGGACAAGAAACAGACAGTGTATGC
ACTTTTTTATAAACCTGATATAGTACCAATTAATGAGAGTGAGTATATATTAATACATTGCAATACCTCAACCATAACA
CAAGCCTGTCCAAAGGTCTCTTTTGACCCAATTCCTATACATTATTGTGCTCCAGCTGGTTATGCGATTCTAAAGTGTA
ATAATAAGACATTCAATGGGACGGGACCATGCCAAAATGTCAGCACAGTACAATGCACACATGGAATTAAGCCAGTAGT
ATCAACTCAACTACTGTTAAATGGTAGCATAGCAGAAGGAGAGATAATAATTAGATCTGAAAATCTGACAAACAATGTT
AAAACAATAATAGTACACCTTAATGAATCTATAGGAATTGTGTGTACAAGACCCGGCAATAATACAAGAAAAAGTATAA
GGATAGGACCAGGACAAGCATTCTATACAAATCACATAATAGGAGATATAAGACAAGCATATTGTAACATTAGTAAACA
AGAATGGAACAAAACTTTAGAAGAGGTGAGAAAAAAATTGCAAGAACACTTCCCAAATAAAACAATAAAATTTAACTCA
TCCTCAGGAGGGGACCTAGAAATTACAACACATAGCTTTAATTGCAGAGGAGAATTTTTCTATTGCAATACATCAAAAC
TATTTAATGATAGTCTAGTAAATGATACAGAAAGTAATTCAACCATCACTATTCCATGCAGAATAAAACAAATTATAAA
CATGTGGCAGGAGGTAGGACGAGCAATGTATGCCCCTCCCATTGCAGGAAACATAACATGTAAATCAAATATCACAGGA
CTACTATTGACACGTGATGGAGGAACAGATAACACAACAGAGATATTCAGACCTGGAGGAGGAAATATGAAGGACAATT
```

GGAGAAGTGAATTATATAAATATAAAGTAGTAGAAATTAAGCCATTGGGAATAGCACCCACTGAAGCAAAAAGGAGAGT
GGTGGAGAGAGAAAAAAGAGCAGTGGGAATAGGAGCTGTGCTCCTTGGGTTCTTGGGAGCAGCAGGAAGCACTATGGGC
GCGGCGTCAATAACGCTGACGGTACAGGCCAGACAACTGTTGTCTGGTATAGTGCAACAGCAAAGCAATTTGCTGAGAG
CTATAGAGGCGCAACAGCATATGTTGCAACTCACAGTCTGGGGCATTAAGCAGCTCCAGACAAGAGTCCTGGCTATAGA
AAGATACCTAAAGGATCAACAGCTCCTAGGACTTTGGGGCTGCTCTGGAAAACTCATCTGCACCACTAATGTGCCTTGG
AACTCCAGTTGGAGCAATAAATCTCAACAAGCTATTTGGGATAACATGACATGGATGCAGTGGGATAGAGAAATTAATA
ATTACACAAACATAATATACCAGTTGCTTGAGGACTCGCAAATCCAGCAGGAACAGAATGAAAAAGATTTATTAGCATT
GGACAAGTGGCAAAATCTGTGGAGTTGGTTTAGCATAACAAATTGGCTATGGTATATAAAAATATTCATAATGATAGTA
GGAGGCTTAATAGGTTTAAGAATAATTTTTGCTGTGCTATCTATAGTAAATAGAGTTAGGCAGGGATACTCACCTTTGT
CGTTTCAGACCCTTACCCCAAACCCGAGGGGACCCGACAGGCTCGGAGAAATCGAAGAAGAAGGTGGAGAGCAAGACAG
AGACAGATCCGTTCGATTAGTGAGCGGATTCTTACCACTTGCCTGGGACGATCTGCGGAGCCTGTGCCTCTTCAGCTAC
CACCGATTGAGAGACTTCATATTCGATTGCAGCGAGGACAGTGGAACTTCTGGGACGCAGCAGTCTCAGGGGACTCCAG
AGGGGTGGGAAGTCCTTAAATATCTGGGAAGCCTTGTGCAGTATTGGGGTCTGGAGCTAAAAAGAGTGCTATTAGTCTG
CTTGATACCCATAGCAATAGCAGTAGCTGAAGGAACAGATAGGATTATTGAATTAGTACTAAGATTTTGTAGAGCTATC
CGCAACATACCTACAAGAGTAAGACAGGGCTGTGAAGCAGCTTTGCTATAA

SEQUENCE I.D. No 14: Du179 Env Protein

```
1       ANFLGKIWPS HKGRPGNFLQ SRPMRVRGIQ RNWPQWWIWG ILGFWMLMIC SGVGNLWVTI
61      YYGVPVWREA KTTLFCASDA KAYDREVHNV WATHACVPTD PNPQEIVMGN VTENFNMWKN
121     DMVDQMHEDI INLWDQSLKP CVKLTPLCVT LKCSTYNGSD TNDMRNCSFN TTTEIRDKKQ
181     TVYALFYKPD IVPINESEYI LIHCNTSTIT QACPKVSFDP IPIHYCAPAG YAILKCNNKT
241     FNGTGPCQNV STVQCTHGIK PVVSTQLLLN GSIAEGEIII RSENLTNNVK TIIVHLNESI
301     GIVCTRPGNN TRKSIRIGPG QAFYTNHIIG DIRQAYCNIS KQEWNKTLEE VRKKLQEHFP
361     NKTIKFNSSS GGDLEITTHS FNCRGEFFYC NTSKLFNDSL VNDTESNSTI TIPCRIKQII
421     NMWQEVGRAM YAPPIAGNIT CKSNITGLLL TRDGGTDNTT EIFRPGGGNM KDNWRSELYK
481     YKVVEIKPLG IAPTEAKRRV VEREKRAVGI GAVLLGFLGA AGSTMGAASI TLTVQARQLL
541     SGIVQQQSNL LRAIEAQQHM LQLTVWGIKQ LQTRVLAIER YLKDQQLLGL WGCSGKLICT
601     TNVPWNSSWS NKSQQAIWDN MTWMQWDREI NNYTNIIYQL LEDSQIQQEQ NEKDLLALDK
661     WQNLWSWFSI TNWLWYIKIF IMIVGGLIGL RIIFAVLSIV NRVRQGYSPL SFQTLTPNPR
721     GPDRLGEIEE EGGEQDRDRS VRLVSGFLPL AWDDLRSLCL FSYHRLRDFI FDCSEDSGTS
781     GTQQSQGTPE GWEVLKYLGS LVQYWGLELK RVLLVCLIPI AIAVAEGTDR IIELVLRFCR
841     AIRNIPTRVR QGCEAALL*
```

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
17 January 2002 (17.01.2002)

PCT

(10) International Publication Number
WO 02/004494 A3

(51) International Patent Classification[7]: C07K 14/16, C12N 7/00, 7/02

(21) International Application Number: PCT/IB01/01208

(22) International Filing Date: 9 July 2001 (09.07.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/216,995	7 July 2000 (07.07.2000)	US
2000/3437	10 July 2000 (10.07.2000)	ZA
2000/4924	15 September 2000 (15.09.2000)	ZA

(71) Applicants *(for all designated States except US)*: **MEDICAL RESEARCH COUNCIL** [ZA/ZA]; Francie van Zijl Drive, Parow Valley, 7500 Cape Town (ZA). **UNIVERSITY OF CAPE TOWN** [ZA/ZA]; Observatory, 7500 Cape Town (ZA). **UNIVERSITY OF NORTH CAROLINA AT CHAPEL HILL** [US/US]; CB 4100, Bynum Hall, Chapel Hill, NC 27599-4100 (US). **ALPHAVAX INCORPORATED** [US/US]; 2 Triangle Drive, Research Triangle Park, NC 27709-0307 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **WILLIAMSON, Carolyn** [ZA/ZA]; University of Cape Town, Observatory, 7500 Cape Town (ZA). **SWANSTROM, Ronald, Ivar** [US/US]; University of North Carolina at Chapel Hill, CB 4100 Bynum Hall, Chapel Hill, NC 27599-4100 (US). **MORRIS, Lynn** [ZA/ZA]; National Institute for Virology, Modderfontein Road, 2131 Sandringham (ZA). **KARIM, Salim, Abdool** [ZA/ZA]; Francie van Zijl Drive, Parow Valley, 7500 Cape Town (ZA). **JOHNSTON, Robert, Edward** [US/US]; University of North Carolina at Chapel Hill, CB 4100, Bynum Hall, Chapel Hill, NC 27599-4100 (US).

(74) Agents: **CLELLAND, Sandra, Luischen** et al.; Spoor and Fisher, P.O. Box 41312, 2024 Craighall (ZA).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report*

(88) Date of publication of the international search report:
13 March 2003

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: PROCESS FOR THE SELECTION OF HIV-1 SUBTYPE C ISOLATES, SELECTED HIV-1 SUBTYPE ISOLATES, THEIR GENES AND MODIFICATIONS AND DERIVATIVES THEREOF

(57) Abstract: The invention provides a process for the selection of HIV-1 subtype (clade) C isolates, selected HIV-1 subtype C isolates, their genes and modifications and derivatives thereof for use in prophylactic and therapeutic vaccines to produce proteins and polypeptides for the purpose of eliciting protection against HIV infection or disease. The process for the selection of HIV subtype isolates comprises the steps of isolating viruses from recently infected subjects; generating a consensus sequence for at least part of at least one HIV gene by identifying the most common codon or amino acid among the isolated viruses; and selecting the isolated virus or viruses with a high sequence identity to the consensus sequence. HIV-1 subtype C isolates, designated Du422, Du 151 and Du 179 (assigned Accession Numbers 01032114, 00072724 and 00072725, respectively, by the European Collection of Cell Cultures) are also provided.

WO 02/004494 A3

INTERNATIONAL SEARCH REPORT

Inten al Application No
PCT/IB 01/01208

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 C07K14/16 C12N7/00 C12N7/02

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 C07K C12N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)
EPO-Internal, EMBL, BIOSIS

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	DE BAAR M.P. ET AL.: "Subtype-specific sequence variation of the HIV type 1 long terminal repeat and primer-binding site" AIDS RES. AND HUMAN RETROVIR., vol. 16, no. 5, 20 March 2000 (2000-03-20), XP002221002 the whole document ---	1-4,27
A	TSCHERNING C. ET AL.: "Differences in chemokine coreceptor usage between genetyc subtypes of HIV-1" VIROLOGY, vol. 241, 1998, pages 181-188, XP002221003 the whole document ---	1-7,9, 11,13-30
	-/--	

[X] Further documents are listed in the continuation of box C.

[ ] Patent family members are listed in annex.

* Special categories of cited documents :

'A' document defining the general state of the art which is not considered to be of particular relevance

'E' earlier document but published on or after the international filing date

'L' document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

'O' document referring to an oral disclosure, use, exhibition or other means

'P' document published prior to the international filing date but later than the priority date claimed

'T' later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

'X' document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

'Y' document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

'&' document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
15 November 2002	02/12/2002

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Galli, I

Form PCT/ISA/210 (second sheet) (July 1992)

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	NOVITSKY VA ET AL: "Molecular cloning and phylogenetic analysis of human immunodeficiency virus type 1 subtype C: a set of 23 full-length clones from Botswana" JOURNAL OF VIROLOGY, THE AMERICAN SOCIETY FOR MICROBIOLOGY, US, vol. 73, no. 5, May 1999 (1999-05), pages 4427-4432, XP002144689 ISSN: 0022-538X -& DATABASE EMBL PROTEINS 'Online! 1 November 1999 (1999-11-01) "Gag polyprotein" retrieved from EMBL Database accession no. Q9WF90 XP002221006 * 95.7% identity with seq. 1 * * 94.5% identity with seq. 5 * * 94.8% identity with seq. 12 * -& DATABASE EMBL PROTEINS 'Online! 1 November 1999 (1999-11-01) "Gag-pol polyprotein" retrieved from EMBL Database accession no. Q9WF89 XP002221007 * aa 457-1139: 86.8% identity with seq. 7 (aa 170-850) * -& DATABASE EMBL DNA 'Online! 11 March 1999 (1999-03-11) "HIV-1 isolate C-96BW04.02" retrieved from EMBL Database accession no. AF110962 XP002221008 * nt 1582-3703: 72% identity with seq. 6 (nt 449-2571) * -& DATABASE EMBL DNA 'Online! 11 March 1999 (1999-03-11) "HIV-1 isolate C-96BW11.04 country Botswana" retrieved from EMBL Database accession no. AF110969 XP002221009 * nt 6152-8143: 89% identity with seq. 10 (nt 591-2579) * -& DATABASE EMBL DNA 'Online! "HIV-1 isolate C-96BW15B03 country Botswana" retrieved from EMBL Database accession no. AF110973 XP002221010 * nt 280-1760: 75% identity with seq. 4 (nt 91-1571) * --- -/--	14-16

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication,where appropriate, of the relevant passages	Relevant to claim No.
X	GAO F ET AL: "Molecular cloning and analysis of functional envelope genes from human immunodeficiency virus type 1 sequence subtypes A through G. The WHO and NIAID networks for HIV isolation and characterization" JOURNAL OF VIROLOGY, THE AMERICAN SOCIETY FOR MICROBIOLOGY, US, vol. 70, no. 3, March 1996 (1996-03), pages 1651-1667, XP002123321 ISSN: 0022-538X -& DATABASE EMBL PROTEINS 'Online! 1 November 1996 (1996-11-01) "Envelope glycoprotein" retrieved from EMBL Database accession no. Q70014 XP002221011 * aa 215-554:85% identity with seq. 3 * * aa 215-348+469-563 about 85% identity with seq. 14 * ---	20
X	LOLE K S ET AL: "FULL-LENGTH HUMAN IMMUNODEFICIENCY VIRUS TYPE 1 GENOMES FROM SUBTYPE C-INFECTED SEROCONVERTERS IN INDIA, WITH EVIDENCE OF INTERSUBTYPE RECOMBINATION" JOURNAL OF VIROLOGY, THE AMERICAN SOCIETY FOR MICROBIOLOGY, US, vol. 73, no. 1, January 1999 (1999-01), pages 152-160, XP002929279 ISSN: 0022-538X -& DATABASE EMBL PROTEINS 'Online! 1 November 1998 (1998-11-01) "envelope protein" retrieved from EMBL Database accession no. O90096 XP002221012 * 83% identity with seq. 9 * * 82% identity with seq. 11 (aa 24-857) --- -/--	22,23

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication,where appropriate, of the relevant passages	Relevant to claim No.
P,X	LEIGH BROWN A.J. ET AL.: "Reduced susceptibility of human immunodeficiency virus type 1 (HIV-1) from patients with primary HIV infection to nonnucleoside reverse transcriptase inhibitors is associated with variation at novel amino acid sites." J. VIROL., vol. 74, no. 22, November 2000 (2000-11), pages 10269-10273, XP002221004 -& DATABASE EMBL PROTEINS 'Online! 1 June 2001 (2001-06-01) "Reverse transcriptase" retrieved from EMBL Database accession no. Q99FC3 XP002221013 * aa25-302: 96.7% identity with seq. 2 * * aa 26-302: 97.5% identity with seq.13 *	18
A	DATABASE EMBL DNA 'Online! 2 January 1996 (1996-01-02) "HIV-1 isolate BU/91/07, envelope" retrieved from EMBL Database accession no. HI1U39249 XP002221014 * nt420-2559: 70% identity with seq. 8 (nt 402-2541) *	13
A	VAN HARMELEN J.H. ET AL.: "A predominantly HIV Type 1 subtype C-restricted epidemic in South African urban populations" AIDS RES. AND HUMAN RETROVIR., vol. 15, no. 4, 1999, pages 395-398, XP002221005	

INTERNATIONAL SEARCH REPORT

ll ıtional application No.
PCT/IB 01/01208

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☒ Claims Nos.: 8,10,12,31
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:
see FURTHER INFORMATION sheet PCT/ISA/210

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

see additional sheet

1. ☐ As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.

2. ☒ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest.

☐ No protest accompanied the payment of additional search fees.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

This International Searching Authority found multiple (groups of) inventions in this international application, as follows:

1. Claims: 1-4

   A process for the selection of HIV subtype isolates.

2. Claims: 5,8,9,16,17,28,31 and partly 24-26

   An HIV-1 subtype C, designated Du422 and its gag,pol,env sequences and consensus sequences.

3. Claims: 6,10,11,12,13,18,19,20,21,29 and partly 24-26

   An HIV-1 subtype C, designated Du151 and its gag,pol,env sequences and consensus sequences.

4. Claims: 7,14,15,22,23,30 and partly 24-26

   An HIV-1 subtype C, designated Du179 and its gag,pol,env sequences and consensus sequences.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

Continuation of Box I.2

Claims Nos.: 8,10,12,31

Claims 8,10,12,31 relate to DNA sequences 1,3 and 5, which are not provided in the application. These claims cannot be searched.

NOTE: Claims 9,11,13-23 relate to sequences 2,4,6,7,8,9,10,11,12,13,14. These Seq. ID refer to the preliminary (obsolete) sequence listing. The claims have been searched using the corresponding sequence IDs 1-11 of the current sequence listing. Claims 24-26 relate to consensus sequences represented by Seq. IDs 12,13 and 14 of the current sequence listing.

The applicant's attention is drawn to the fact that claims, or parts of claims, relating to inventions in respect of which no international search report has been established need not be the subject of an international preliminary examination (Rule 66.1(e) PCT). The applicant is advised that the EPO policy when acting as an International Preliminary Examining Authority is normally not to carry out a preliminary examination on matter which has not been searched. This is the case irrespective of whether or not the claims are amended following receipt of the search report or during any Chapter II procedure.